THIS WEEK

Combating schizophrenia

Research has revealed daunting complexities in the psychiatric condition, but also new routes towards diagnosis and treatment.

Some people hear voices that are the products only of processes in their brains. These hallucinations can lead to persistent delusions that someone is plotting against them or urging them to harm others. When a person acts on those delusions, headline-grabbing tragedy can ensue, usually involving someone close to the protagonist.

SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

Violence is not a symptom of schizophrenia — only a tiny proportion of sufferers with the condition are homicidal. Yet these incidents dominate the media coverage of the disease. The reality of schizophrenia is much more complex. Hallucinations are one of several symptoms, others of which — cognitive dysfunction, loss of motivation and of social engagement — are much less amenable to medication, and are often more damaging to the ability of those with schizophrenia to function. In recent years it has been increasingly appreciated that this collection of symptoms, which typically first fully manifest in early adulthood, represents a late stage of the illness, and that the illness itself may perhaps turn out to be a collection of syndromes, rather than a single condition.

Motivated by the undue stigma and by the recent advances reported in our own pages and elsewhere, *Nature* this week examines the state of our understanding of schizophrenia, and how researchers can hope to make progress in an entangled landscape of innate and environmental influences. The image on this issue's cover and in the logo that links the associated content is a piece of art by a schizophrenia sufferers. It is one of many compiled by NARSAD, a US charity based in New York that spends significant public donations on psychiatric health research. The image reflects a world of confusion and distorted reality — but not a 'split personality', which is a mythical symptom of the condition, and leads to a misleading metaphorical use of the word 'schizophrenic' that those involved with the condition perpetually seek to eradicate.

Globally, somewhere between 0.5% and 1% of the population will experience schizophrenia in their lifetime. It is not the most prevalent psychiatric condition — clinical depression, aka unipolar disorder, is the most common. But analyses by the World Health Organization and others show that it has a disproportionate social cost. Moreover, owing to the few high-profile tragic episodes, it is probably the psychiatric condition with the most stigma attached — a stigma that is persistent, as those who have fought it have discovered time and again (see page 163).

In some societies, schizophrenia is not recognized as a disease — the response to its symptoms can be very dependent on culture, as can the degree of support that sufferers receive. Where it is recognized and treatment is given, it poses some of the greatest challenges in medical care. Drugs that treat the delusional symptoms have been around for decades, and have helped many sufferers enormously. But these drugs have not advanced in effectiveness, and several manufacturers have now backed out of this arena (see page 158). Cognitive and behavioural therapies have also made progress, but they suffer from a lack of coherent research and due priority (see page 165). In the current round of development of clinical diagnostics of the disease, the very definition of schizophrenia is under debate (see page 168).

One of the key diagnostic and scientific challenges is to distinguish the early phases of this condition, using all the means of biological, behavioural and cognitive analyses at our disposal. As with most diseases, the earlier the diagnosis, the greater the ability to mitigate the disease's impact or to defeat it. Recent research indicates that it may be possible to diagnose an 'at-risk' condition at the earliest phases of schizophrenia — for that reason as well as others, studies of adolescent brain development and behaviour have become a crucial research pathway (see pages 154 and 187). More generally, studies of the brain's morphology and circuitry (see page 194), and of the ways in which genetic and environmental influence interact (see page 203), are poised to make substantive progress.

We are in a decade in which fundamental science and clinical and social research can make an enormous impact on mental health (see *Nature* **463,** 9; 2010). Those who have schizophrenia, and their families and colleagues who are also affected by the condition, depend on research to deliver new diagnostics and treatments. The rapid progress in science is daunting in the complexities that it has revealed. But it is yielding much for sufferers to hope for and new paths for researchers to follow. Funders please note. ∎

States or the union

California climate initiative moves decisively forward, providing a glimmer of hope.

Before the Democrats took control of Congress in 2006 and the White House in 2008, US climate strategists focused much of their energy at state and local level. Their goal was threefold: to test new climate policies, to reduce greenhouse-gas emissions where possible and to increase pressure on businesses and policy-makers to craft a viable national compromise. The third strand of that strategy failed this year, making the successes attained in the other two all the more important. From this perspective, the vital environmental vote of last week's mid-term elections was in California, which forcefully rejected the ballot aiming to scuttle the state's pioneering climate law, Proposition 23.

Sixty-one per cent of California voters weighed in against Proposition 23, which put forward short-sighted arguments that the state's drive to reduce greenhouse-gas emissions would hobble an already

struggling economy. Current law requires the state to cut greenhouse-gas emissions by 25% on 1990 levels by 2020, and the California Air Resources Board is bringing in regulations to make that happen, including a cap-and-trade system that would allow businesses to decide where to make emissions reductions. California has long led the United States on environmental policy, and this is exactly the kind of action that will expand and could, in time, pave the way for a more comprehensive approach on climate in Washington DC.

California has already been joined by seven other states in the country's west, and four Canadian provinces, in the Western Climate Initiative. Farther east, six states and one Canadian province have signed the Midwestern Greenhouse Gas Reduction Accord. These two programmes aim to cut greenhouse-gas emissions by 15% and 20%, respectively, on 2005 levels by 2020. Carbon trading has already begun under the first such programme, the Regional Greenhouse Gas Initiative, under which ten northeastern states have committed to reducing current levels of emissions by 10% by 2018.

In total, 23 US states and five Canadian provinces have begun their own climate initiatives, independent of their countries' federal governments. The Washington DC-based World Resources Institute calculates that these initiatives cover half the US population and one-third of its greenhouse-gas emissions. In Canada, the numbers are even more impressive: more than three-quarters of the population and half the greenhouse-gas emissions are covered. Twenty-eight states have standards on renewable energy, and countless other efforts are under way that would encourage energy savings and greenhouse-gas reductions at the state and local level.

And do not count out the federal government just yet. The administration of President Barack Obama is preparing to roll out new greenhouse-gas regulations next year and has a variety of other tools at its disposal. The administration has ordered the federal government, the largest energy consumer in the nation, to reduce emissions by 28% on 2008 levels by 2020. The federal government — and the defence department in particular — also has enormous purchasing power. If used wisely, that could help drive green technologies to market.

Clearly, these limited efforts are not sufficient. Concerned citizens in the United States and around the world are right to be disappointed in the lack of leadership and vision in Congress. Barring a sharp reversal on their campaign rhetoric, it seems that the situation will only get worse when Republicans take control of the House of Representatives next year. But many of these Republicans hail from those states that are quietly embarking on their own programmes to combat global warming. These aren't just feel-good measures, either. They are confidence builders.

> "As policy labs of long standing, US states can demonstrate the power of 'bottom-up' leadership."

It's possible that, globally, climate change is simply too complex a problem for a comprehensive top-down solution. There are too many interests at stake, too many losers with loud voices. And although most Americans clearly accept the reality of global warming, most people simply don't care that much, particularly when economic woes loom large. That's not a recipe for success in the power corridors of Washington. But California and other US states are policy labs of long standing, and as such will continue to demonstrate the power of 'bottom-up' leadership. ∎

Worth waiting for

A wise report on genetic screening from the Leopoldina has been 350 years in the making.

This week, the Leopoldina, Germany's national academy of sciences in Halle, is due to publish a historic report. The publication analyses how advances will affect the field of predictive genetic diagnosis, and calls for changes in a law that has confused science and ideology. It is also notable for being the first report to come from the Leopoldina in its role as a national academy — three and a half centuries after it was set up.

The law, which came into force in February, had been debated for nearly a decade. The legislation was intended to protect the population from the possible abuse of genetics, but instead presents an unreasonable threat to the health of individuals with treatable genetic disorders.

One problem is that the law ranks patient confidentiality above a doctor's responsibility to the health of that patient's relatives. It rules, for example, that genetic data collected for diagnosis should be destroyed after ten years, even though the guidelines of the German Chamber of Physicians say that such data should be retained for at least 30 years to provide for the health of the next generation. The law also misunderstands details of science. To name but one instance, it redefines neonatal screening for genetic disease — used routinely for decades to identify 12 treatable genetic disorders by chemical, not DNA, analysis — as 'genetic screening'. This means that a doctor with expertise in genetic counselling, rather than a midwife, must take blood for the test. This complicates processes, and in small rural hospitals where such expertise is not available some doctors are reportedly simply choosing not to screen.

These problems have their roots in a cultural fear of sharing medical data — a legacy of the Nazi era. With Germany's federalized health-care system, patients can move between doctors of their choosing without any of their medical history following them. This situation makes it impossible to carry out optimal population screening for genetic diseases, and will restrict the future health value of new technologies.

The Leopoldina, too, is embroiled in the legacies of history — and has seen a lot of history flow past since it was founded as a scientific academy in 1652. In 2007, federal research minister Annette Schavan unilaterally declared that the Leopoldina would become Germany's national academy. In doing so, she put an end to years of acrimonious debate about which, if any, of the country's seven regional academies should be elevated to this status. That ruffled feathers. Germany's federal structure was designed as a core element of its 1949 constitution, to prevent any centralization of power and thus to ensure that a fascist regime could never take control again. Individual states were given wide powers and have inevitably become protective of them. All this left Germany without a national academy to provide, among other things, authoritative and independent scientific advice for policy-making. With its first report, the Leopoldina has already shown the value of such an institution.

The academy established a committee of 17 scientists and legal experts to analyse how new genomic technologies and other advances are set to affect predictive genetic diagnosis, and how Germany can better prepare itself for the opportunities and ethical challenges that will follow. The resulting report, *Predictive Genetic Diagnostics as an Instrument of Disease Prevention*, calls for the creation of national centres of competence to overcome these barriers. There are legal hurdles to this, but they can, and should, be overcome.

The Leopoldina's report responds to a shadow of history. Politicians should follow its recommendations and change the law to prevent further damage. It is time for Germany to see its past in the appropriate historical context, and to ensure that its psychological legacies do not inadvertently harm the innocent today. And the Leopoldina? From its base in what was once an important centre of chemical industry in the former East Germany, it will be able to direct a little bit of history, not just watch it flow by. ∎

↻ **NATURE.COM**
To comment online, click on Editorials at:
go.nature.com/xhunqv
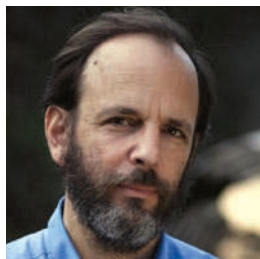
Double trouble? To throw cash at science is a mistake

Politicians who seek economic recovery should look beyond the budget of the National Science Foundation, argues **Daniel Sarewitz**.

Several weeks before the recent US election, I testified before the House of Representatives' Committee on Science and Technology. The topic of the hearing was the 'science of science policy'. In my allotted five minutes, I explained that science investments should be assessed not just in terms of the amount of money spent, but in terms of their capacity to achieve desired social outcomes, such as improved public health or environmental quality.

When it came to questions from the members of Congress attending the hearing — all two of them — Representative Vern Ehlers (Republican, Michigan) ignored my testimony and explained how things were done. Noting that a Republican-led Congress had doubled the budget of the National Institutes of Health (NIH) between 1998 and 2003, Ehlers said that he hoped a new Republican majority in the House would follow suit and this time double the budget of the National Science Foundation (NSF).

Ehlers, who retires this year, holds a PhD in physics and has served on the science committee since 1994. So he as much as anyone in Congress should know what makes good science policy. Apparently, it is merely a mastery of the two-times table. President Barack Obama also backs an NSF doubling.

With the United States mired in debt and deficit, and fiscally conservative Republicans now in charge of the House, many government programmes, including research and development (R&D), will stagnate or be slashed in the next two years. But I predict that the new Republican majority will ally itself with the president to significantly boost funds for the NSF.

Can this be plausible? Haven't the Republicans been at war with science, attacking climate-change research and blocking work on embryonic stem cells? Maybe so, but for nearly a decade, Republican lawmakers have also taken the lead in attempts to double the NSF budget. Former House speaker Newt Gingrich, a fiscal conservative who led the 1994 Republican takeover of Congress, has said that he wants to triple the NSF budget.

Why will money flow to the NSF? When times are hard and the political and economic news is grim, throwing money at basic research makes for feel-good politics that plays well in both parties and the media. It also responds to the appeals of scientific and political elites who are obsessed with the idea that the nation is losing its scientific and technological leadership in a globalizing world. Nor is it expensive. The NSF accounts for only 7% of non-defence R&D spending.

Boosting funds for basic research is also safer politics than actually tackling a national problem. When Congress doubled the NIH budget, Republicans and Democrats alike claimed that they were improving health care by supporting science. But witness what happened when Democrats tried to address the failings of the nation's medical system directly by passing health-care reform legislation earlier this year. They provoked a powerful political backlash that contributed to last week's electoral defeats.

Unlike almost any other area of government investment, the public accepts that benefits from basic research are unpredictable and long term. Politicians thus get to fob off accountability on the scientific community, whose only responsibility is to ensure that funding goes to the 'best' science. It's no surprise, then, that in many ways the NIH and NSF are the government's crown jewels, extolled by Republicans and Democrats alike because they give away billions of dollars to scientists efficiently, transparently and with minimal scandal or controversy.

So, doubling the NSF budget makes good politics. But what about good policy? Although few would question the value of a robust basic-science enterprise, we just don't know how marginal increases in basic-research funding affect a nation's capacity to solve social and economic problems. On the other hand, decades of research on the links between science and innovation in areas ranging from jet engines to medicines show that basic research best contributes to economic and social goals when targeted at areas that can directly benefit from additional fundamental knowledge.

> **DOUBLING THE NSF BUDGET MAKES GOOD POLITICS. BUT WHAT ABOUT GOOD POLICY?**

If Congress wanted to allocate scarce new R&D resources wisely during a protracted period of budgetary austerity, it wouldn't adopt a doubling strategy, but would instead take a more surgical approach to set priorities. It would focus investments where links between science and application are well established, to deliver short- to medium-term benefits. Alternative energy, for example, offers many technological options where basic research can improve performance. And if Congress is going to cut R&D, it could look towards research areas that, despite huge, long-term investment, have shown weak links to desired social outcomes. Much basic research on cancer and on climate change fall into this category.

I know that such an approach would be fiercely opposed by the leading voices of the scientific community, who will never abandon the long-falsified belief that basic research is most valuable to society as a bottom-up enterprise driven only by inherent scientific interest. Moreover, the political attractiveness of generic basic research funded by agencies such as the NSF could be one of the few things that Republicans and Democrats agree on in the next two years. But we shouldn't fool ourselves into thinking that the most politically expedient way to support science is also the most socially beneficial. ∎

Daniel Sarewitz, *co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, is based in Washington DC.*
e-mail: dsarewitz@gmail.com

M. TEMCHINE

Catalyst makes dyeing greener

A 'green' polymer could become even greener thanks to a process that eliminates the need for an environmentally damaging dyeing step.

Poly-lactic acid is a polyester manufactured from molecules derived from plants. The dyeing process typically uses large volumes of water and strong alkali, and generates sulphur-contaminated waste water. Patrick McGowan at the University of Leeds, UK, and his colleagues use an aluminium-based catalyst that polymerizes lactic acid or lactides and also carries the dye molecule. This allows the polymerization and dyeing steps to be combined into one, and obviates the need for additional harmful chemicals.

This 'DyeCat' method also reduces production costs and could be used by the textile industry, the authors say.
Angew. Chem. Int. Edn **doi:10.1002/anie.201004920 (2010)**

A. BSHARY

Chasing off biters benefits others

Behaviour that seemingly contributes to the public good can evolve as a by-product of self-serving actions.

Andrea and Redouan Bshary at the University of Neuchâtel in Switzerland studied a type of blenny fish (*Plagiotremus rhinorhynchus*) that stealthily bites scalefin anthias (*Pseudanthias squamipinnis*; pictured) from behind.

Some victims chase the biting blenny after an attack. The authors used Plexiglass plates to mimic chasing or fleeing anthias in the lab, and found that chasing acts as a self-serving punishment, prompting blennies to pursue other individuals in future attacks. In the field, chased blennies were seen biting other prey species, suggesting that chasing also serves a public good.

Moreover, the researchers found that blennies can distinguish between chasers and those that flee or 'free-ride'. They say that free-riders might make themselves easy targets for the blennies, thereby favouring the chasing behaviour.
Curr. Biol. **doi:10.1016/j.cub.2010.10.027 (2010)**

Neutral molecules whirl around

Particle accelerators typically whip charged particles around using magnetic and electric fields, enabling physicists to investigate high-energy collisions. Peter Zieger of the Fritz Haber Institute in Berlin and his colleagues have

built a prototype molecular synchrotron for neutral molecules, a key step towards studying low-energy collisions.

Using specially designed electric fields, the team managed to keep ammonia molecules centred as they whizzed around a 0.5-metre-diameter ring (pictured). Varying the strength of the

electric fields allowed the group to keep the molecules grouped in 'bunches'. Bunching up the molecules should increase the sensitivity of future collision studies.
Phys. Rev. Lett. **105,** 173001 **(2010)**

Trimming brain connections

Immune cells called microglia help to protect the brain after an injury. They may also be involved in pruning the connections, or synapses, between neurons — a key process in learning and

memory formation.

Using electron microscopy, Marie-Ève Tremblay, Rebecca Lowery and Ania Majewska at the University of Rochester in New York imaged mouse brain slices and reconstructed the interactions between microglia and synapses in three dimensions. Most of the microglia were directly adjacent to the synapses, and in particular to dendritic spines — neuronal structures — that were small and were often pruned away later on.

The authors also subjected mice to shifts in visual experience using light deprivation and re-exposure. After this regime, some

S. PUTZKE, FHI

microglia more often displayed features that allow them to engulf synaptic elements, suggesting a role for these cells in trimming back synapses after new experiences.

PLoS Biol. 8, e1000527 (2010)

MATERIALS SCIENCE

Film bends with light

A polymer film that has been sandwiched between two Teflon sheets curls when hit with ultraviolet (UV) light, a result of the film's highly ordered three-dimensional (3D) molecular structure. Such 3D ordering over macroscopic length scales could aid in the design of advanced materials and devices.

Takuzo Aida and Takanori Fukushima at the RIKEN Advanced Science Institute in Saitama, Japan, and their colleagues designed a brush-shaped polymer bearing multiple photoresponsive azobenzene groups. They then hot pressed the polymer film between two Teflon sheets that were stretched along one axis. Spectroscopic measurements revealed that, during hot pressing, the Teflon sheets imposed a certain molecular order on the film, such that the concerted motion of the photoresponsive units caused the film to curl in response to UV only when the drawing directions of the Teflon sheets had been parallel to each other.

Science 330, 808–811 (2010)

AGRICULTURE

Sterile moths fight resistance

Releasing sterile moths into cotton fields keeps levels of the dreaded pink bollworm moth (*Pectinophora gossypiella*) to a minimum — a strategy that could prevent the pest from becoming resistant to insecticide-producing crops.

Pink bollworms gradually evolve resistance to the Cry1Ac toxin made by genetically modified cotton, especially if farmers fail to maintain nearby refuges of non-toxic crops.

Computer simulations developed by Bruce Tabashnik at the University of Arizona in Tucson and his colleagues suggested that releasing sterile bollworm moths would decrease the bollworm population and lessen the chance that two Cry1Ac-resistant moths would mate and propagate the trait. The model predicted that after more than 20 years, Cry1Ac resistance would not emerge, even in the absence of refuges.

A four-year field experiment on modified and conventional cotton across the state of Arizona backed up the computer model: between 2006 and 2009, pink bollworm infestation rates plummeted by 99.9%.

Nature Biotechnol. doi:10.1038/nbt.1704 (2010)

For a longer story on this research, see go.nature.com/gsurwm

ASTRONOMY

Wave patterns in Saturn's big ring

The structures in the outer edge of Saturn's most massive ring, the B ring, appear to be oscillating in at least three rotating wave patterns. This may be helping to create the complex variation seen in the ring's behaviour.

Until recently, researchers thought that the region was dominated solely by the gravitational influence of Saturn's moon Mimas. Joseph Spitale and Carolyn Porco of the Space Science Institute in Boulder, Colorado, pored over thousands of images of the B ring region (pictured) taken by NASA's Cassini spacecraft during its initial four-year exploration of the planet. The duo discovered the wave patterns, which arise spontaneously owing partly to the high density in the ring. They also found evidence for the existence of small moons trapped in the region.

The findings offer insight into similar oscillations in other astronomical systems, such as spiral galaxies, and protoplanetary disks orbiting nearby stars.

Astron. J. 140, 1747–1757 (2010)

ZOOLOGY

No sons for snake that shuns sex

For the most part, snakes reproduce sexually, with males carrying two Z sex chromosomes and females one Z and one W. Warren Booth of North Carolina State University in Raleigh and his colleagues have identified an unprecedented natural occurrence of WW females born to a single female *Boa constrictor imperator*.

By analysing the DNA of the female and two of her broods that were all-female, the researchers ruled out sexual reproduction with cohabiting male boas. However, the offspring were not clones of their mother, and the authors suggest that they were the result of automictic parthenogenesis. In this form of reproduction, the egg's nucleus fuses with another cell structure in the egg that carries half of the mother's chromosomes.

This is the first evidence of multiple, viable and non-experimentally induced WW females in a vertebrate, the authors say.

Biol. Lett doi:10.1098/rsbl.2010.0793 (2010)

IMMUNOLOGY

It's not you. It's your organelles.

Transplanted cells that contain their new host's nuclear DNA could still be rejected by the individual's immune system. This is because of slight differences between the cells and the host — in the genomes of the energy-producing organelles called mitochondria.

Jun-Ichi Hayashi at the University of Tsukuba in Japan, Noriko Toyama-Sorimachi at the National Center for Global Health and Medicine in Tokyo and their colleagues compared tumour cells containing nuclear DNA from one mouse strain and mitochondria from either the same or a different strain.
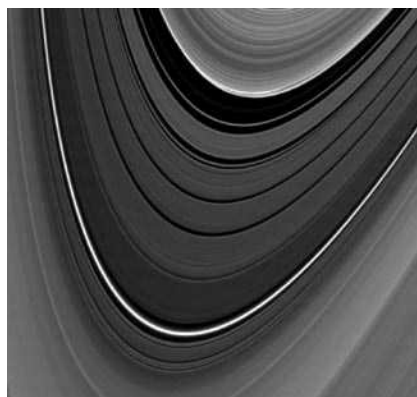
Despite just a 0.5% difference between the two mitochondrial genomes, cells containing the foreign organelles did not form tumours when transplanted into the strain of mice that the nuclear DNA came from. However, the hybrid cells grew at the same rate as normal tumour cells in mice lacking certain types of immune cell.

Tissues made from a person's stem cells could face rejection when transplanted back into him, the researchers say, because mitochondrial genomes tend to accumulate mutations.

J. Exp. Med. doi:10.1084/jem.20092296 (2010)

NASA/JPL/SPACE SCIENCE INSTITUTE

SEVEN DAYS

The news in brief

US–India ties

On a state visit to India, US President Barack Obama announced on 8 November that his administration would ease restrictions on the export of technologies to India's defence and space-related agencies. He added that reforms to export controls would promote a greater partnership between the two nations. India has been subject to US sanctions on high-tech exports since its nuclear-weapons tests in 1998.

Oil-spill faults

A White House-appointed commission investigating the explosion at BP's Deepwater Horizon oil rig in the Gulf of Mexico published preliminary findings and held public hearings in Washington DC on 8 and 9 November. The findings laid out a list of faults leading up to the blowout, including human error and technical failures, but said that there was "no evidence at this time to suggest that there was a conscious decision to sacrifice safety concerns to save money". The commission's report is scheduled to be released in January 2011.

Election fallout

After Republican success in the US midterm elections last week, a divided Congress — coupled with public concern over government spending — could herald a period of budget-cutting for US research. See page 146 for more.

Education costs

Higher education in science and medicine looks set to become much more expensive in England. On 3 November, the UK government announced that universities will be able to charge students fees of up to £6,000 (US$9,700)

Cometary close encounter

Bright jets of carbon dioxide exploding from a rough surface featured in most images of comet Hartley 2 taken by NASA's EPOXI spacecraft during its 4 November flyby (see *Nature* **467,** 1013–1014; 2010). But the midsection of the comet's oddly shaped 2.2-kilometre-long nucleus is mysteriously smooth. The probe — which in 2005 visited and shot a projectile into comet Tempel 1 as part of the Deep Impact mission — came within 700 kilometres of Hartley's nucleus, the fifth to be photographed since an armada of probes flew past comet Halley in 1986.

a year, and in some cases up to £9,000. Those taking degrees in laboratory-based subjects are likely to face the biggest hikes. The move is a key part of the government's plan to overhaul university funding after it decided in October to cut the higher-education budget by £2.9 billion over four years. The current cap on student fees is £3,290.

Polar-bear status

The US Department of the Interior has until 23 December to explain its 2008 decision during the administration of president George W. Bush to list the polar bear as a 'threatened' — but not 'endangered' — species, a federal judge ruled on 4 November. The order follows a lawsuit challenging the 'threatened' listing, brought by environmental campaigners including the Center for Biological Diversity, based in Tucson, Arizona. They hope that a change of listing to 'endangered' may mean that the Endangered Species Act can be used to protect the bears' sea-ice habitats by enforcing the regulation of greenhouse-gas emissions.

Detention protest

Nineteen physicists have written to the French Physical Society to protest against the detention of Adlène Hicheur, a 33-year-old French-Algerian researcher who has been held on suspicion of terrorism for more than a year without a trial. Hicheur was a postdoc at the Swiss Federal Institute of Technology in Lausanne (EPFL) before his arrest in France on 8 October 2009 for alleged ties to Al-Qaeda in the Islamic Maghreb — the North African branch of the terrorist organization. The physicists warn that the "length and the arbitrariness of the procedure" could damage Hicheur's career. See go.nature.com/xakh8r for more.

Biotech squeeze

Biogen Idec — a prominent US biotechnology firm that is one of the few first-generation biotechs not to have been consumed by big pharma — said on 3 November that it would make major cuts to save some US$300 million a year. The company, based in Weston, Massachusetts, will lay off around 650 employees, or 13% of its workforce, and close three facilities — two in Massachusetts and one in San Diego, California. To sharpen its focus, Biogen will jettison 11 research programmes, including its cardiovascular and oncology projects, in favour of its core area,

neurology. The firm has seven drugs in or near late-stage clinical trials, including four for treating multiple sclerosis.

Arthritis drug

A new class of treatment for the autoimmune condition rheumatoid arthritis has passed a phase III clinical trial of 611 patients, announced its developer, multinational drug giant Pfizer, this week. The oral pill, tasocitinib, inhibits a kinase protein involved in stimulating immune cells. It is one of the farthest advanced of several kinase inhibitors jockeying to replace current biological drugs against rheumatoid arthritis, which must be injected. Pfizer will not seek marketing approval until trials pitting tasocitinib against other medicines are completed next year.

Carbon burial

The Dutch government has cancelled a plan by oil company Shell to inject 400,000 tonnes of compressed carbon dioxide annually into ageing natural-gas beds below Barendrecht, the Netherlands. The much-delayed project was stopped owing partly to a "complete lack of local support", minister of economic affairs Maxime Verhagen told parliamentarians on 4 November. The decision was not surprising, as Barendrecht residents have opposed the scheme for years;

their protests (**pictured**) made the project an emblem for opposition to carbon capture and storage (see *Nature* **463,** 871–873; 2010).

Fertilizer bid block

Canada's government said on 3 November that it would not approve a US$39-billion bid by mining giant BHP Billiton for the fertilizer company PotashCorp, based in Saskatoon, Canada. The bid was driven by an anticipated growth in demand for — and thus profits from — phosphates, nitrogen and potash. But industry minister Tony Clement said that the deal would not provide net benefit to Canada. BHP, which is based in Melbourne, Australia, has 30 days to revise its offer and change government minds.

Biotech tax credits

A complete list of the 2,923 small US biotechnology companies sharing more than 4,600 awards from a $1-billion tax-credits programme was released last week. The Therapeutic Discovery Project

programme, established by the health-care reform bill passed earlier this year, awards credits and grants to firms with no more than 250 employees (see *Nature* **465,** 854–855; 2010). Big winners include Theravance, based in South San Francisco, which received nearly $2.7 million for 11 projects, and Arisaph in Boston, Massachusetts, with almost $2.8 million for 12 projects. See go.nature. com/tyd2ef for more.

RESEARCH

Collider concludes

The Large Hadron Collider (LHC) has finished smashing protons for the year. The particle accelerator at CERN, Europe's premier high-energy physics laboratory near Geneva, Switzerland, ended its 2010 run on 4 November. It collided particles at 7 teraelectronvolts: a power seven times greater than that of its closest rival, the Tevatron at Fermilab in Batavia, Illinois. But in terms of collision numbers, the amount of data gathered this year matched what the Tevatron can do in only a week. The LHC is now focusing on colliding lead ions.

Nuclear sharing

As part of a major defence-sharing agreement, the United Kingdom and France have agreed to build a joint test

11–12 NOVEMBER
Discussions on how to stabilize the world's supply of rare-earth elements have been urged for the G20 meeting in Seoul.
www.g20.org/index.aspx

13–17 NOVEMBER
Treating depression with deep-brain stimulation and a campaign for more money for brain research are on the agenda for the Society for Neuroscience's meeting in San Diego.
go.nature.com/lysp8t

16 NOVEMBER
The Royal Society in London releases a report mapping how science is practised around the world, and how the patterns are changing.
go.nature.com/iv3bqu

17–18 NOVEMBER
In Cadarache, France, ITER's council will discuss measures to cut costs at the international fusion project.
go.nature.com/bokzva

facility for nuclear weapons research. Called EPURE, the facility will X-ray dummy bombs as they explode to understand the effects of ageing on nuclear warheads. EPURE will be located at the Valduc Centre for Nuclear Studies near Dijon, the main French weapons lab, and will be developed with the UK Atomic Weapons Establishment at Aldermaston. Neither side gave a cost estimate, although money-saving was said to be the motivation for the facility, which is expected to be operational from 2015. See go.nature.com/womdnf for more.
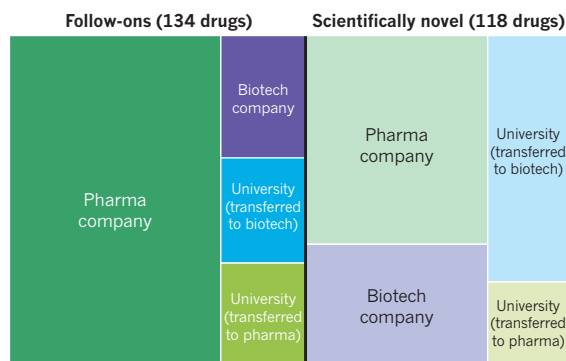
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

BUSINESS WATCH

Pharmaceutical companies discovered 58% of 252 drugs approved by the US Food and Drug Administration during 1998–2007 (see chart). But two-thirds of pharma drugs approved were follow-on compounds. The industry produced a smaller proportion of innovative drugs — universities and biotech firms made 56% of such discoveries. The analysis is from a study by Robert Kneller at the University of Tokyo, published on 29 October (R. Kneller *Nature Rev. Drug Discov.* **9,** 867–882; 2010).

ORIGINS OF NEW DRUGS

Universities and biotech companies accounted for more than half of the scientifically innovative drugs* approved by the US Food and Drug Administration from 1998 to 2007.

Follow-ons (134 drugs)		Scientifically novel (118 drugs)	
Pharma company	Biotech company	Pharma company	University (transferred to biotech)
	University (transferred to biotech)		
	University (transferred to pharma)	Biotech company	University (transferred to pharma)

*Those with new mechanisms of action, or that were first-in-class

NEWS IN FOCUS

More than a billion people rely on drinking water fed by Himalayan glaciers such as Khumbu in Nepal. This supply could be at risk unless glacial retreat is stopped.
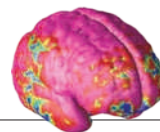
J. BENDIKSEN/MAGNUM PHOTOS

CLIMATE CHANGE

Measuring the meltdown

With global warming hitting the Tibetan plateau hard, scientists gather to plan an international research campaign to understand and mitigate changes at the 'third pole'.

BY JANE QIU IN KATHMANDU, NEPAL

Cold, remote and threatened by global warming: the description applies not only to the North and South Poles, but also to a region of more than five million square kilometres, centred on the Tibetan plateau and the Himalayas, that researchers call the third pole (see *Nature* **454,** 393–396, 2008). After the Arctic and the Antarctic, the region has Earth's largest store of ice, in more than 46,000 glaciers and vast expanses of permafrost. Yet it is much less studied than its high-latitude counterparts, even though many more lives depend on it.

The third pole is also known as Asia's water tower, because its glaciers feed the continent's largest rivers, which sustain 1.5 billion people across ten countries. Those glaciers are melting fast, filling lakes that can overflow and flood valleys. Yet little is known about how climate change is unfolding there. To attempt to rectify this, Third Pole Environment (TPE), an international programme led by the Chinese Academy of Sciences' Institute of Tibetan Plateau Research (ITP) in Beijing, held its second workshop last month in Kathmandu. Researchers in the region laid plans to fill the knowledge gap, and discussed findings that add to the urgency.

"Everyone is doing important work across the region, but it is unclear how they fit together," says Yao Tandong, director of the ITP and chairman of the TPE science committee, which

helped to organize the workshop. "The only way forward is for the international community to work together to assess the risks associated with climate change," he says.

As the region's population booms, the top priority of researchers is to understand the status and fate of glaciers that are a vital source of drinking water. Last year, a claim in the 2007 report of the Intergovernmental Panel on Climate Change (IPCC) that Himalayan glaciers could disappear by 2035 turned out to be an error (see *Nature* **463,** 276–277; 2010). But participants at the workshop argued that the IPCC's broader concern about the rapid loss of Himalayan glacier ice was correct. "There is no doubt that many of the glaciers in the region ▶

are retreating fast," says Baldev Raj Arora, former director of the Wadia Institute of Himalayan Geology in Dehradun, India.

But it is not clear exactly how fast, or how this will affect water resources, because there is no glacier inventory for the entire region. Taken alone, satellite studies offer only a rough estimate of glaciated area, and remoteness, high altitudes and harsh weather conditions hamper measurement from the ground.

The evidence that is available is telling. Using a combination of satellite and ground measurements, a team led by Liu Shiyin, a glaciologist at the Chinese Academy of Sciences' Cold and Arid Regions Environmental and Engineering Research Institute in Lanzhou, has just completed the second Chinese national glacier inventory, documenting some 24,300 glaciers and recording characteristics such as their locations, lengths and surface areas. It shows that the total surface area of glaciers has decreased by 17% and that many have disappeared since the last inventory began, roughly 30 years ago.

To get a better estimate of such changes, researchers have also measured the ice volume and mass balance of representative glaciers in various parts of the Himalayas. Such arduous studies, often at altitudes above 5,000 metres, show that "the impact of climate change on some Himalayan glaciers is much worse than previously thought", says Tian Lide, a glaciologist at the ITP. The expansive Kangwure glacier on the northern slope of Mount Xixiabangma in southern Tibet, for instance, has lost nearly half of its ice since the 1970s, and its average thickness has decreased by 7.5 metres.

Most of the glaciers in the Indian Himalayas that have been studied in detail are also losing mass, says Arora. Across the third pole, "regional anomalies exist, but the balance of evidence certainly points towards a trend of fast retreat", agrees Lonnie Thompson, a glaciologist at Ohio State University in Columbus and a co-chairman of the TPE science committee.

CARBON TOLL

One cause of the retreat is the growing amount of sooty 'black carbon' made by fossil-fuel and biomass burning. Xu Baiqing, an environment scientist at the ITP, measured 50 years' worth of black-carbon levels in ice cores from five glaciers in various parts of the Himalayas, and found increased emissions since the 1990s, coinciding with rapid industrial growth in the

region. Angela Marinoni, a climate scientist at the Institute of Atmospheric Sciences and Climate in Bologna, Italy, and her colleagues found high concentrations of aerosols, including black carbon, above 5,000 metres in the Nepalese Himalayas, which caused significant atmospheric warming. They calculate that deposition of black carbon could increase snow and ice melting of a typical Himalayan glacier by 12–34%, by reducing its ability to reflect light.

As a consequence, glacial lakes are getting larger and more numerous, causing more floods. A study led by Yongwei Sheng, an ecologist at the University of California, Los Angeles, shows that the area of such lakes on the plateau has increased by 26% since the 1970s, with a devastating effect on surrounding pastures. Outbursts of glacial lakes have caused more than 40 floods in the Himalayas since the 1950s, and more are likely in the coming decades, says Pradeep Mool, a remote-sensing specialist at the International Centre for Integrated Mountain Development (ICIMOD) in Kathmandu. An ICIMOD survey listed more than 20,200 glacial lakes in the region; two hundred are "potentially dangerous" and need close monitoring and an

KAZAKHSTAN · MONGOLIA · Kunlan · TIBETAN PLATEAU · CHINA · Himalayas · NEPAL · BHUTAN · INDIA · MYANMAR

early warning system, says Mool.

So far, scientists trying to forecast the glaciers' future have had little to go on. For one thing, says Thompson, "glaciers respond to climate differently depending on their size, altitude distribution, surface area, debris cover and valley characteristics". And little is known about how climate is changing across the third pole.

Yang Kun, a climate scientist at the ITP, found that many satellite-based measurements of Earth's radiation budget — the balance of incoming solar radiation and outgoing heat

— did not work well at the third pole's high elevations, because instruments are typically calibrated and verified against ground-truth data from the lowlands. They can be corrected using field measurements, but across the region, only 16 weather stations lie above 5,000 metres.

Investigators can't rely on climate models, either. Using data from the only weather station at 8,000 metres, on the South Col pass between mounts Everest and Lhotse, Kenichi Ueno, a climate scientist at the University of Tsukuba, Japan, showed that global climate models do not predict moisture and radiation flux well at such high altitude, especially in hot weather or monsoon season. "If you want to know how climate affects glaciers, such details are crucial," he says. "It's extremely important to have more high-altitude observations across the region."

A JOINT ENDEAVOUR

The TPE science committee will soon draft a research programme to document the effects of climate change on glaciers, permafrost, water resources, biodiversity and people. The plan, to be finalized by autumn 2011, will call for joint expeditions to the Himalayas and the Tibetan plateau, and for multidisciplinary research stations across the region, to cover key geological areas and climate regimes as well as important river and lake basins. Once costs have been defined, the committee will seek support from national and international funding agencies.

The most important part of the plan is a common data repository, says Volker Mosbrugger, director of Senckenberg World of Biodiversity, a coalition of research institutes and museums in Frankfurt, Germany, and another co-chairman of the TPE science committee. But national rivalries may stand in the way, especially when the shared data concern water resources. "Whether there can be a central database up and running will determine whether the programme can move beyond its rhetoric," says Gregory Greenwood, director of the Mountain Research Initiative at the University of Bern. "This will be a great challenge."

The committee will draft a policy to be negotiated between countries interested in the programme, letting scientists share information but leave out data deemed politically sensitive. "Without working together and pooling data from across the third pole," says Yao, "comprehensive understanding of climate impact and feedback mechanisms would be impossible." ∎

MORE ONLINE

IN PICTURES
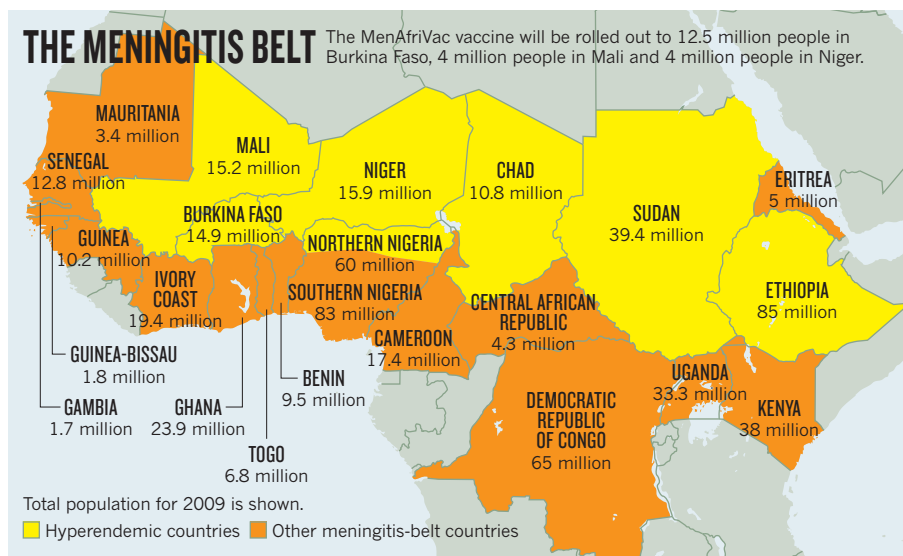Oil spill cruise finds field of dead coral after the Deepwater Horizon disaster go.nature.com/oxw17x
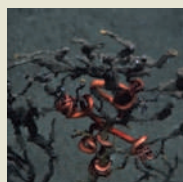
DAILY NEWS
● Sterile moths wipe out cotton pest go.nature.com/9wxufg
● Stunted development in cystic fibrosis begins at birth go.nature.com/cdhybu
● Mayans converted wetlands into farmland go.nature.com/bwavz3

OPINION
If heat is needed to speed up the earliest biochemical reactions, water may be life's only crucible, argues Philip Ball. go.nature.com/abrnwe

SOURCE: PATH

THE MENINGITIS BELT

The MenAfriVac vaccine will be rolled out to 12.5 million people in Burkina Faso, 4 million people in Mali and 4 million people in Niger.

MAURITANIA
3.4 million

SENEGAL
12.8 million

MALI
15.2 million

NIGER
15.9 million

CHAD
10.8 million

ERITREA
5 million

GUINEA
10.2 million

BURKINA FASO
14.9 million

NORTHERN NIGERIA
60 million

SUDAN
39.4 million

IVORY COAST
19.4 million

SOUTHERN NIGERIA
83 million

CENTRAL AFRICAN REPUBLIC
4.3 million

ETHIOPIA
85 million

GUINEA-BISSAU
1.8 million

CAMEROON
17.4 million

UGANDA
33.3 million

GAMBIA
1.7 million

GHANA
23.9 million

BENIN
9.5 million

DEMOCRATIC REPUBLIC OF CONGO
65 million

KENYA
38 million

TOGO
6.8 million

Total population for 2009 is shown.
🟨 Hyperendemic countries 🟧 Other meningitis-belt countries

PUBLIC HEALTH

Vaccine offers meningitis hope

First affordable and effective weapon against killer meningococcal meningitis A rolled out in Africa.

BY DECLAN BUTLER

From Senegal to Ethiopia, December heralds the arrival of intense epidemics across Africa's 'meningitis belt' (see map). Killing thousands, and leaving many more with a host of after-effects such as brain damage and deafness, the scourge of meningitis A terrifies communities.

This year will be different. Millions will receive a new vaccine, MenAfriVac, that promises protection against the meningococcal bacterium *Neisseria meningitidis*. It is the culmination of ten years' work by an international consortium to develop a vaccine at a price low enough for massive use in Africa: just US$0.40 a dose. "MenAfriVac is a fantastic initiative," says Andrew Riordan, a meningitis expert at Alder Hey Children's NHS Foundation Trust, in Liverpool, UK. "For the first time, we may be able to prevent these huge epidemics."

The Meningitis Vaccine Project (MVP), led by the World Health Organization (WHO) and PATH, a non-profit body based in Seattle, Washington, was born in 2001 after a particularly bad epidemic in 1996–97 caused 250,000 cases and 25,000 deaths (see 'Epidemic cycle'). Commercial manufacturers in developed countries could not produce the vaccine at such a low target price, according to Marc LaForce, director of the MVP. So the consortium did the research

itself, and contracted the Serum Institute of India in Pune to make the vaccine. The entire research and development cost of the project was just $70 million — five to ten times less than typical vaccines. LaForce hopes that the MenAfriVac model can be applied successfully to other vaccines.

During next month's campaign, backed by the WHO and the United Nations Children's Fund (UNICEF), the government of Burkina Faso will vaccinate everyone aged 1–29 — the group hit hardest by the disease, numbering 12.5 million people. Mali and Niger will each vaccinate 4 million people in the same age bracket.

Meningitis A epidemics cause fewer cases and deaths in Africa than AIDS or malaria, but this masks its huge social and economic toll in those

SOURCE: MVP

EPIDEMIC CYCLE

Africa's meningitis epidemics usually take place in irregular cycles every 5–12 years.

countries. "When the epidemic arrives, the entire community shuts down," says LaForce. The disease — which infects the meninges, the membranes surrounding the brain and spinal cord — begins with mild symptoms of stiff neck, high fever, confusion and headache, but can kill within 48 hours. Of those infected with meningococcal meningitis A, 5–10% die and 10–20% of survivors are left with severe disabilities.

Until now, the only weapon against meningitis A in Africa was a polysaccharide vaccine that offered protection for a very short time. Using it as a preventative vaccine would have required repeated vaccination of the population — a logistically and financially impossible approach. So it was used only to vaccinate people after an epidemic was under way, often too late to have an impact. By contrast, MenAfriVac is a longer-lasting conjugate vaccine, in which an antigen is coupled to a protein to provoke a stronger immune response. LaForce says that clinical-trial subjects who received the vaccine more than three years ago are still protected.

The previous polysaccharide vaccine also did not work well in children, whereas MenAfriVac is approved for children aged one year and over, and trials are under way for children as young as one month, with results expected next year. MenAfriVac has the added advantage of stopping people from becoming carriers that spread the disease, protecting unvaccinated people through 'herd immunity'.

In June, the WHO decided that the vaccine met international standards of safety and efficacy, and results from phase III trials have been submitted for journal publication. The roll out will be accompanied by a 'pharmacovigilance' system for reporting adverse events.

Several international groups, including the US Centers for Disease Control and Prevention in Atlanta, Georgia, and the Norwegian Institute of Public Health in Oslo, will run surveillance programmes to prove the vaccine's effectiveness to donors and governments.
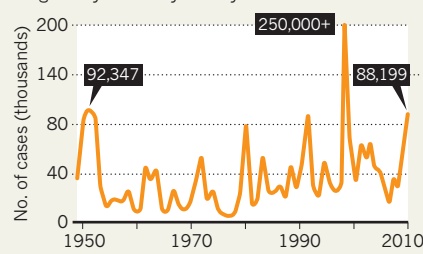
Yet millions may miss out. Some $475 million is needed to expand coverage to other countries in the meningitis belt, says Marie-Pierre Preziosi, an official at the WHO's Department of Immunization, Vaccines and Biologicals in Geneva. Raising the money will be challenging, she says, particularly because the GAVI Alliance, the main sponsor of vaccination in low-income countries, is facing a financial crunch (see *Nature* **464,** 338; 2010)

Ultimately, public-health officials would like to deploy several conjugate vaccines that protect against not just meningococcal type A but also the other subgroups in the region that can cause smaller meningitis epidemics.

For now, researchers are satisfied that an affordable meningitis vaccine will at last be available in Africa. As Andrew Pollard, a meningitis researcher at the Jenner Institute at the University of Oxford, UK, says: "The roll out of MenAfriVac in Burkina Faso is a triumph for this unique partnership." ■

MENTAL HEALTH

China tackles surge in mental illness

Psychological examinations to be added to selection procedure for government officials.

BY DAVID CYRANOSKI

In a country where the government likes to emphasize the harmony of its society and the well-being of its people, China's announcement on psychological examinations came as a surprise. A top official in the country's ruling Communist party said last month that the examinations will be introduced into the selection procedure for senior government officers and leaders of national industries. According to state media, the move was in part a response to a string of suicides by officials, many of whom suffered from depression.

It is the highest-level recognition yet of a problem that is widely affecting China's people, from the central arms of government to distant rural villages. According to a survey published last year, 17.5% of Chinese have some form of mental illness, one of the highest rates in the world (M. R. Phillips *et al. Lancet* **373,** 2041–2053; 2009). In comparison, a World Health Organization survey found that the United States had the highest prevalence of mental illness in the world, with 26.4% of the population thought to have a diagnosable disorder in any given year (The WHO World Mental Health Survey Consortium *J. Am. Med. Assoc.* **291,** 2581–2590; 2004).

In China, some blame the pressures of the social transformation accompanying the country's economic expansion; others argue that changes in the way mental illness is diagnosed might explain the rise. But with mental illness now firmly on the list of national health concerns, researchers, drug companies and medical practitioners are quickly developing new initiatives to tackle the issue.

IDENTIFYING RISK

Last year, the Chinese science ministry invested a total of 40 million yuan (US$6 million) in ten two-year research projects that target issues such as prevention of suicide and depression, identifying early warning signs of the risk of developing mental illness, and dealing with occupational stress. For example, Wanjun Guo, a psychiatrist at the

Mental Health Center of West China Hospital in Sichuan, recently got 20 million yuan to "organize a survey on mental disorders in the 'post Wenchuan earthquake' region". Raymond Chan, a psychologist and deputy director of the Chinese Academy of Sciences Key Laboratory of Mental Health in Beijing, hopes that the funding will continue. He is seeking support for a databank of 'endophenotypes': characteristics that are thought to reflect a

MENTAL-HEALTH MARKET
Eli Lilly, the leading provider of mental-health drugs in China, has seen rapid growth in its sales of antidepressant and antipsychotic drugs.

genetic disposition to mental diseases, such as eye-tracking dysfunction or memory failure in people with schizophrenia. "This will be a low cost, quick diagnostic tool," says Chan. "China probably needs it more than anywhere else."

China is poorly equipped to bring new findings into the clinic, however. Yanling He, a psychiatrist at Shanghai Mental Health Center, says that there are only 20,000 psychiatrists in China: 1.5 for each 100,000 people, a tenth of the ratio in the United States. Many are poorly qualified, and they often deal with problems that would be handled by social workers or clinical psychologists elsewhere. Patchy distribution is an even greater worry, adds He. Although there are 200 psychiatrists at the Shanghai Mental Health Center alone, the only practitioner in Tibet recently left, she says.

International collaborations are helping to strengthen China's mental-health infrastructure. Psychiatrists from the University of Melbourne in Australia have been working with the Chinese University of Hong Kong and the Peking University Institute of Mental Health

to train people and establish community-based health programmes, for example. The success of the collaboration has convinced China's health ministry to fund a scale-up to more than 100 sites, says University of Melbourne psychiatrist Chee Ng, who helped to lead the effort. And Elise Snyder, a New York-based psychoanalyst, is organizing the China American Psychoanalytic Alliance, which for the past two years has been training some of China's first psychotherapists through Skype conversations with US-based practitioners.

DIAGNOSING CHANGE

A change in government policy on medical insurance in 2005 also gave Chinese patients access to modern antidepressants and antipsychotics from abroad. "I see increased investment in mental-health infrastructure and wider coverage of essential treatment for patients," says Peter Salzmann, vice-president for marketing in China for pharmaceutical company Eli Lilly (see 'Mental-health market').

China's ministry of health has brought its mental-health guidelines and resources in line with international standards. The latest version of the *Chinese Classification of Mental Disorders* (*CCMD*), published in 2001, made Chinese definitions "nearly a replica" of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) issued by the American Psychiatric Association, according to Sing Lee, a psychiatrist at the Chinese University of Hong Kong.

But Lee worries that these changes may have led to the apparent boom in mental-health disorders. The latest *CCMD* follows the *DSM* in noting that depression can be diagnosed after just two weeks of symptoms — the previous duration recognized in China was three months, says Lee. "That will surely lead to high prevalence," he says. "It also increases the risk of medicalization of normal sadness like grief."

He adds that making discontent and unhappiness medical problems by focusing on psychological or psychiatric factors could distract attention from underlying social ills that might be to blame. Several of the government officials who committed suicide became depressed when they were investigated for corruption — a stark reminder that more psychiatrists and antidepressants alone might not be enough to cure China's mental-health woes. ∎

SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

C. OWEN/AP

House speaker-in-waiting Republican John Boehner says limited government is top of the agenda.

US MIDTERM ELECTIONS

US science faces a squeeze

Power shift in Congress paves way for Republicans to constrict government spending.

BY RICHARD MONASTERSKY,
JEFF TOLLEFSON, MEREDITH WADMAN
AND EUGENIE SAMUEL REICH

Restricted budgets; investigations of climate research; hearings on ocean science; debates about NASA's future. The US elections last week, which gave Republicans a commanding control of the House of Representatives, could throw a series of challenges at science over the next two years.

Reacting to the moribund economy and concerns about governmental excess, voters handed Republicans their largest House majority since 1946. Republicans gained at least 60 seats in the House and wrested control from the Democrats, who had taken over the lower legislative body just four years before. But Democrats retained a slight majority in the Senate, the upper legislative branch, thereby limiting the ability of Republicans to make many of the sweeping changes they had pledged during the lead up to the elections.

After the election, Charles Vest, president of the National Academy of Engineering based in Washington DC, said that it was important for US scientists to respect the system and work with the new Congress on issues such as research funding and science education. "I hope we'll find some champions there," he says.

In the past, both parties have generally supported increases in funding for basic research, with budget levels rising or falling depending more on the state of the economy than on the party controlling the purse strings. But many of the Republicans elected last week — especially those belonging to the 'Tea Party' — ran on platforms to reduce government spending. John Boehner (Republican, Ohio), who is expected to become the House speaker, or leader, said: "It's pretty clear the American people want us to do something about cutting spending here in Washington."

Boehner and fellow Republicans in the current Congress issued a 'Pledge to America' in September that promised to return government spending to the levels of fiscal year 2008. If applied across the board, that would mean a 7% cut from 2010 spending levels for research and development not related to defence, according to an analysis by the American Association for the Advancement of Science (AAAS).

President Barack Obama, too, plans cuts, but it is unclear how they will affect science. The budget for the current fiscal year is still being finalized (see 'Unfinished business'). His budget office in June indicated that 2012 funding for non-security agencies would drop 5% below levels previously projected for that year, but science might avoid the worst. Obama said last week: "I don't think we should be cutting back on research and development."

Supporters of science will appeal to both parties to spare research from the most painful cuts. "It's incumbent upon scientists to be able to articulate the research that they're doing and why this research is important," says Joanne Carney, director of the Center for Science, Technology and Congress at the AAAS.

The election means new leadership — and new scepticism about environmental issues — in several key committees that oversee research spending and strategy. The Republicans will select chairpersons in early December, and senior legislators are now jockeying for key posts.

Leadership of the House Committee on Science and Technology is expected to go to 87-year-old Ralph Hall of Texas, who is the Republicans' top, or 'ranking', member on the committee. Hall issued a statement after the election saying he sees a role for science policy in driving innovation and the US economy, which researchers could take as a positive sign. But Hall also outlined his intention to scrutinize the Obama administration's policies in areas including climate change and energy research.

Climate scientists could face harsher treatment in the House Committee on Oversight and Government Reform, expected to be led by Darrell Issa of California. Issa has said he will look into the recent debates about climate science sparked by the release of e-mails from the Climatic Research Unit at the University of East Anglia, UK, a year ago.

On the Committee on Energy and Commerce, ranking Republican Joe Barton of Texas is battling to retake the gavel of the committee he chaired before the Democratic takeover in 2007. Barton achieved notoriety in climate circles five years ago when he investigated the famous 'hockey stick' graph depicting global average temperatures over the past millennium.

Barton vows to investigate efforts by the Environmental Protection Agency (EPA) to craft greenhouse-gas regulations under the Clean Air Act. The Democratic health-care overhaul is also on his radar. But party rules may hinder his plans. Barton has reached the official limit of six years as the top Republican on the committee, and needs a waiver to serve as chair. If Barton's extension request is unsuccessful,

ELECTIONS

�»NATURE.COM
More on the midterm elections at:
go.nature.com/ midterm2010

Fred Upton of Michigan is the front-runner to be the next chair. Upton has previously endorsed the science of global warming and is viewed as a moderate on environmental issues. Nonetheless, he has recently attacked the EPA for pushing a "regulatory train wreck" targeting everything from greenhouse gases and energy efficiency to conventional air pollutants.

Upton and others are unlikely to mount a direct attack on the EPA's authority — granted by the Supreme Court — to regulate greenhouse gases. Any legal restriction would fall prey to Obama's presidential veto. But Republicans could effectively block the agency by inserting 'riders' into appropriations bills that would prevent the EPA from spending money on specific regulatory programmes.

"The EPA has a job to do, and we need to fight and make sure we protect that ability," says Gene Karpinski, president of the League of Conservation Voters, a non-profit environmental organization based in Washington DC.

Both conservationists and scientists will be watching to see who heads the Committee on Natural Resources. The top contender is Doc Hastings of Washington, who has pledged to challenge the president on such matters as offshore drilling and plans to oversee use of marine resources. Hastings has said that Obama's ocean policy will "lock up vast portions of our oceans through an irrational zoning process".

THE 'LAME-DUCK' SESSION
Unfinished business

When the existing Congress reconvenes on 15 November for its final weeks, members will try to plough through a long list of tasks, including passing the appropriations bills that set government spending levels for the current fiscal year.

President Barack Obama requested a US$1-billion increase to the National Institutes of Health (NIH) budget in February, which received support in congressional committee votes this summer. But that increase could be trimmed as lawmakers attempt to control discretionary spending.

Congressional subcommittees also voted to give increases to the National Science Foundation and the National Oceanic and Atmospheric Administration, but these budgets await final approval by Congress.

Science boosters will also be pushing the Senate to reauthorize the America COMPETES Act, which was enacted in 2007 and expired in October. The legislation put science agencies on a path to double their budgets by 2017. Although Congress would be hard-pressed to provide all of the promised increases, reauthorizing the act would signal a commitment to protecting science during lean budget years.

Republicans will also target Obama's plans to cancel NASA's Constellation programme, which is developing the hardware to send astronauts to the Moon and Mars. In February, Obama proposed shifting the job of developing the capsule and rocket for manned spaceflight to the commercial space industry. Hall has opposed that plan, as has Frank Wolf (Republican, Virginia), who is likely to take over the appropriations subcommittee that oversees NASA's budget.

To make good on their pledges, however, Republican leaders must unite the disparate parts of their caucus, which could be difficult. Many of the incoming legislators from the Tea Party won by challenging the Republican establishment. David Goldston, the chief of staff for the House science committee from 2001 to 2006, when the Republicans last controlled the House, says: "The big story is the extent of division within the Republican ranks." ∎

GRANTS

Funding crisis hits US ageing research

Shortfalls hamper scientists' efforts to address a predicted epidemic of age-related diseases.

BY MEREDITH WADMAN

Karen Duff, an Alzheimer's disease researcher at Columbia University in New York, recently submitted three grant applications to the National Institutes of Health (NIH). In the covering letters, she requested that they not be directed into the pool of applications competing for money from the National Institute on Aging (NIA).

"I don't put any grants into NIA now because their funding line is so low it's almost impossible to get funded," says Duff.

Duff's situation reflects a crisis that is gripping researchers who are funded by the US$1.1-billion ageing institute in Bethesda, Maryland. In 2010, a researcher submitting a grant application for any single deadline had only an 8% chance of winning funding (see graph). That number may soon dip even lower as grants tied to economic-stimulus funding begin to expire and a climate of austerity descends on the United States.

The dismal odds of winning an NIA-funded grant "threaten the viability of ageing research" says Richard Hodes, the NIA's director. "If we are less able to fund research — or are perceived to be less able — that will actually drive young and emerging investigators to fields other than ageing. That would be a catastrophe at a time when such research is critically important."

Although the funding situation is tight all around for NIH-supported investigators, the NIA is in an exceptional predicament. In recent years, it has made big commitments to costly clinical trials and large group projects, even as both the number and the average cost of investigator-initiated grants submitted to the institute have surged. Responding last month to a deluge of concern from researchers, Hodes posted an open letter to the NIA-funded community.

"We are cutting back on the very research that we need to keep our ageing population healthy."
Karen Duff

OVER THE HILL
Competition for grants from the National Institute on Aging has become increasingly brutal.
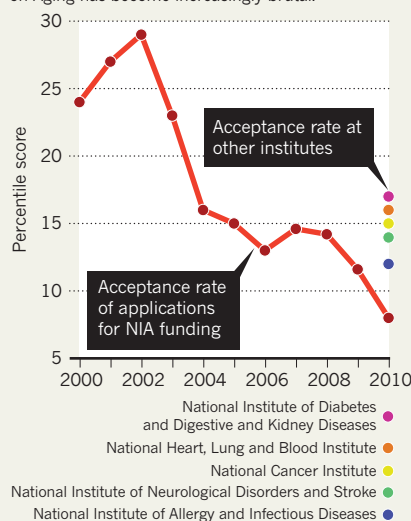
"We at NIA recognize and empathize with the struggle that our constrained funding creates for the research community," he wrote. "It is vital that we do everything we can to sustain the momentum of investigator-generated research in this successful and vibrant field."

Advocates for research into Alzheimer's disease are adding to the chorus, noting that the toll of the condition will rise drastically with the greying of the global population. The NIA funds nearly three-quarters of the NIH's $469-million investment in Alzheimer's disease.

"Our government is ignoring what is likely to become the single greatest threat to the health of Americans," begins a recent opinion piece in *The New York Times*. The co-authors include Nobel laureate and neuroscientist Stanley Prusiner, and Sandra Day O'Connor, a retired US Supreme Court justice whose husband died of complications related to Alzheimer's disease a year ago. The article calls for the US Congress to pass legislation that would increase research funding for Alzheimer's to $2 billion annually and establish a government office that would coordinate efforts to combat the disease, which is expected to affect 13.5 million US citizens by 2050.

As both the US and the global populations age, the prevalence of chronic diseases such

as cancer, heart disease and diabetes will also grow, along with neurodegenerative ailments including Parkinson's disease. The NIA deals with age-related aspects of all of these. "We are cutting back on the very research that we need to keep our ageing population healthy," says Duff.

By contrast, in Europe, "the field of ageing is prioritized. I would certainly not suggest that we are anywhere close to a crisis" in terms of research, says Alan Walker, director of the New Dynamics of Ageing Programme at the University of Sheffield, UK, and head of the European Research Area in Ageing, a network of national research councils and ministries working together to support ageing research.

To battle the dismal success rate of its grant applicants, the NIA is trying to spread what wealth there is. Since 2004 it has, on average, made competitive awards at amounts that are 18% below the budgets that applicants requested. It has established an expert advisory panel to help to evaluate clinical-trial proposals, and it has made it tougher for scientists to submit grants that request more than $500,000 in annual direct costs. The results are apparent: in the past year, the average amount requested by scientists on a standard R01 grant application has fallen by $30,000.

The brutal competition is already taking its toll. Chris Conrad, a colleague of Duff's at Columbia, last month left a tenure-track position for a job in industry. Conrad, 37, had applied for more than ten NIA-supported grants for his work on the genetics of Alzheimer's disease since he came to Columbia in 2006. Not a single one was funded. "They are going to lose a lot of new people by having these really hard funding rates," he says.

Even established researchers are feeling the pinch. Vera Novak, a neuroscientist at Beth Israel Deaconess Medical Center in Boston, Massachusetts, and an NIA grant recipient since 2001, is studying older patients with diabetes to look at its effects on blood circulation in the brain's smallest vessels. Her stimulus-funded NIA grant runs out in July 2011, and she is applying for a standard grant to allow her study — which has enrolled 70 patients out of its target of 120 — to continue after that. She is feeling anxious about her prospects of winning that new grant. "I've never seen a situation as difficult as it is now," says Novak. ∎

CELL BIOLOGY

There will be blood

Direct conversion of cell types could offer safer, simpler treatments than stem cells.

BY EWEN CALLAWAY

In a feat of cellular alchemy, human skin cells have been transformed into blood cells without first being sent through a primordial, stem-cell-like state. For the developers of patient-specific cell therapies, the result could be safer and simpler than induced pluripotent stem (iPS) cells — reprogrammed adult cells that can differentiate into many cell types.

Published in *Nature*[1], the study follows work earlier this year showing that fibroblast cells from mouse skin can be transformed into neurons[2] and heart muscle[3]. However, it is the first study to accomplish direct reprogramming with human cells, and the first to create progenitor cells — in this case for blood. "It takes us a step along the line to believing that you can produce anything from almost anything," says Ian Wilmut, director of the Medical Research Council Centre for Regenerative Medicine in Edinburgh, UK, who was not involved in the study.

Mickie Bhatia, a stem-cell researcher at McMaster University in Hamilton, Canada, and his colleagues infected skin cells with a virus that inserted the *OCT4* gene, then they grew the cells in a soup of immune-system stimulating proteins called cytokines. The gene's product, the OCT4 protein, is one of a handful of factors used to transform fibroblasts into iPS cells, but Bhatia's team found no evidence that the blood progenitor cells they made went through an embryonic state. For instance, the cells did not cause mice to develop teratomas — tumours that are characteristic of pluripotent cells, making iPS cells less attractive as a therapeutic option. The progenitors did, however, produce all three classes of blood cells — white, red and platelets — all of which were functional. The red blood cells also produced the adult form of haemoglobin, whereas iPS-cell-derived blood cells make the fetal form. "This is the most encouraging result we've seen for using blood cells for cell-replacement therapy," says Bhatia.

Converted cells aren't without their drawbacks, though. Unlike iPS and embryonic stem cells, they cannot easily multiply in the lab, so producing the large quantities needed for screening drugs, for example, could be difficult, says Wilmut. It is also too early to tell whether they will be as good as the real thing when inside a person, says George Daley, a stem-cell biologist at Children's Hospital Boston in Massachusetts. Transplanting the cells into humans is still years away, says Bhatia. "The clinical side is going to be a lot of work." ∎

1. Szabo, E. *et al. Nature* advance online publication doi:10.1038/nature09591 (2010).
2. Vierbuchen, T. *et al. Nature* **463,** 1035–1041 (2010).
3. Ieda, M. *et al. Cell* **142,** 375–386 (2010).

↻ NATURE.COM
For a longer version of this story, see:
go.nature.com/ekshcj

CORRECTION

The News Feature 'The telescope that ate astronomy' (*Nature* **467,** 1028–1030; 2010) wrongly said that a revised budgetary plan for the James Webb Space Telescope could come as early as 2 November. In fact, the plan is not likely to be unveiled before the president's budget request in February 2011.

THE MAKING OF A TROUBLED MIND
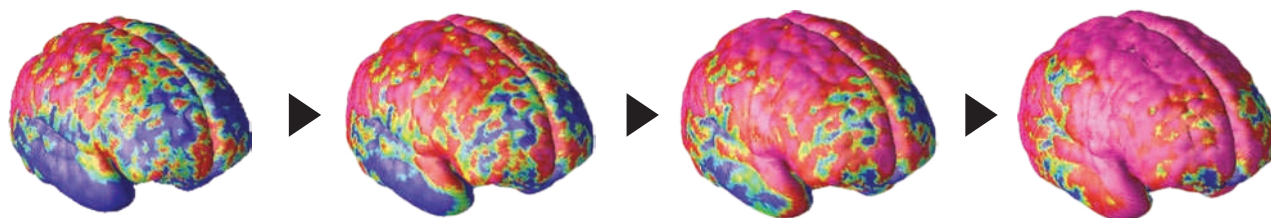
Schizophrenia appears during adolescence.
But where does one begin and the other end?

BY DAVID DOBBS

Rachel had just given birth to her third child when she became overwhelmed by the noise on the obstetrics ward, grew sharply paranoid about her sister, and in short order descended into her first schizophrenic episode. She was 28. Although it was only then that she started hearing voices — those of her family, distant screams, messages from spaceships — she and her psychiatrist came to see that there had been whisperings of this long before.

As a child — bright, but awkward both socially and physically — Rachel tended to keep to herself. She crammed her drawings full of the sort of elaborate fractal detail often seen in the work of psychotic artists. In her teenage years, some of her difficulties worsened. Acutely sensitive to noise, she was aware of the refrigerator cycling off and on, footfalls from the apartment next door, the traffic outside. Only in retrospect did any of these peculiarities seem ominous.

As Rachel's psychiatrist Robert Freedman explains in his book *The Madness Within Us,* where he writes about Rachel, that's how it is with the early flickers of paranoia, confusion, hypersensitivity and hallucination in people who develop schizophrenia. They often emerge exactly when adolescence is throwing the body and brain for a loop, and years before the disease manifests itself fully. "The problem with early symptoms," says Freedman, who is chair of psychiatry at the University of Colorado, Denver, and editor-in-chief of the *American Journal of Psychiatry*, "is

During adolescence grey matter is lost (pink), a process that may speed up in early-onset schizophrenia — hinting at the origins of the disease.

that they're not very specific. At a time when thinking, emotion and behaviour change a lot anyway, these early indicators are very hard to separate from normality."

Even so, this overlap of schizophrenia's early signs with the hallmarks of adolescence has made this period a beacon to researchers. Over the past 20 years, studies have shown that the adolescent brain undergoes major developmental changes. Autopsy and imaging studies, for instance, have revealed that during childhood and adolescence the brain routinely prunes away vast numbers of synapses — the junctions between neurons across which electrical signals flow — and that this pruning seems to go on longer and farther in people with schizophrenia. Other work has shown that adolescence brings major upgrades to the neural networks that generate powers of judgement, cognition and behavioural control — building new circuits, remodelling old ones and discarding some altogether. The idea that schizophrenia arises from miscues or shoddy work in this complicated and delicate project has sparked a huge variety of research. Many basic neuroscientists are trying to work out what goes wrong on genetic, cellular, circuit and systems levels. Meanwhile, at the level of diagnostic practice, some researchers argue that subtle symptoms can not only be distinguished from normal adolescence, but can provide a reliable indicator of future disease.

In a pattern all too familiar to students of schizophrenia, none of these efforts has revealed the secret of this fiendishly complex disorder. One leading researcher, David Lewis, at the University of Pittsburgh's Western Psychiatric Institute and Clinic in Pennsylvania, has spent the past two decades exploring schizophrenia's developmental roots. Yet even Lewis says it's still too soon to know whether any given line of study, no matter how promising, is homing in on the schizophrenia puzzle's most essential component, if such a thing exists.

"It's more like getting a much better picture of one part of the elephant," he says, referring to the old parable of blind men collectively describing an elephant's nature by individually feeling its different parts. "I think it's working. When I talk with other researchers working other ideas, I'm encouraged that I'm onto something important, and even more encouraged that we all seem to be feeling our way around the same animal."

TEEN BRAIN

By many accounts, Lewis is running one of the more comprehensive and sustained attempts to explore normal and pre-schizophrenic adolescent brains. He is taking what neurologist and depression expert Helen Mayberg at Emory University School of Medicine in Atlanta, Georgia, describes as "one of the smartest, most creative and most promising angles I know of on schizophrenia".

Lewis has focused on a particular circuit in the dorsolateral prefrontal cortex (DLPFC), a multilayered region that is crucial to tying the threads of experience, memory, thought and emotion into a coherent, consistent view of the world (see 'Cellular culprits in schizophrenia'). The DLPFC builds and refines much of its complicated circuitry during childhood and adolescence, responding to both genes and experience. Much of Lewis's work examines the relationship, during this developmental period, between two types of cell: pyramidal neurons, which span several layers of the cortex; and chandelier cells, which sit near the base of the pyramidal cells. His concentration on them is one of those stubborn scientific projects that seems to produce little until, suddenly, it produces a lot.

Pyramidal cells, so called because their central bodies are triangular, generate much of the complex electrical signalling that takes place in the DLPFC and in the prefrontal cortex as a whole. Their effectiveness in this task depends heavily on the richness of branching in their long, tree-like forms. Multiple postmortem studies, including some by Lewis, show that in adults with schizophrenia, these pyramidal cells have smaller cell bodies and fewer of the protrusions called dendritic spines that receive input from synapses[1]. In the pyramidal cells in layer 3 of the DLPFC,

which communicates extensively with other cortical regions and is key to powers of working memory that often falter in schizophrenia, Lewis found the dendritic spines reduced in number by about a quarter.

Most developmental neuroscientists suspected that this sparse branching resulted from aberrant synaptic pruning during adolescence. Pruning is a sort of clean-up job conventionally thought to eliminate weak synapses and leave strong ones. In schizophrenia, it was suspected, the pruning process hacked away indiscriminately, knocking out strong synapses along with weak ones.

In 2008, Lewis's team found evidence that argued against this idea in the brains of normal monkeys, whose PFCs develop along timelines comparable to those of humans[2]. The group found that the great majority of synapses in the layer 3 pyramidal neurons were functionally mature before pruning started. This left few immature or weak synapses to trim away. If this holds with healthy humans as well, Lewis argued, then schizophrenia cannot arise from a failure to select and prune away weak synapses, for there is not much to select. This and other discoveries led Lewis to offer an alternative hypothesis: that pyramidal cells in individuals with schizophrenia had weaker synapses before pruning ever began[1,3].
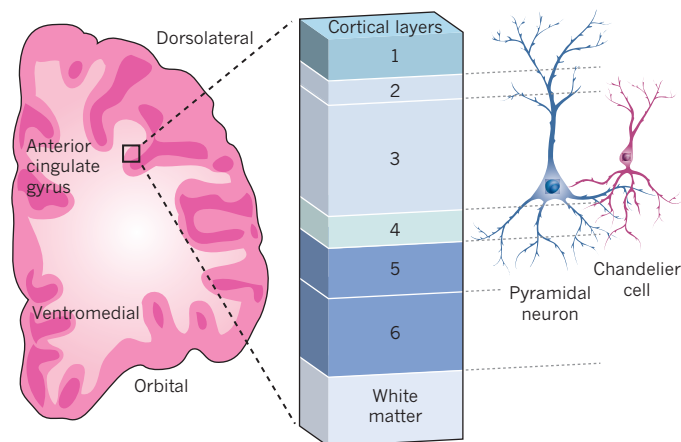
"We have to draw in dotted lines here," says Lewis. "But our suspicion is that early on, when someone destined for schizophrenia has an excess of synapses, the quality of the synapses doesn't matter so much, and the person does okay. Then later, when the pruning starts, the problems slowly become apparent, because they've lost their reserves."

Lewis wanted to trace the story further back than that. What might be stunting the development of pyramidal cells in the first place? He suspected chandelier cells. Over the years, Lewis's group and others have built an increasingly strong case against the ornately branched chandelier cells, which seemed to communicate only with pyramidal cells. Lewis's autopsy work, for instance, had found that chandelier cells took huge hits in schizophrenia[4], with some proteins in their synapses reduced by about 40% — a sign that these junctions were not working normally.

For many years, everyone assumed that chandelier cells only had an inhibitory role: when 'talking' to pyramidal neurons, they said only 'calm down'. Then, about five years ago, a team led by Gábor Tamás at the University of Szeged in Hungary found that chandelier cells also have an excitatory role, sometimes shouting 'fire up'[5]. Groups led by Tamás, Lewis and others have all done work showing that the chandelier cells seem to be key contributors to the complex activity of their pyramidal neighbours. Last year, for instance, Karl Deisseroth's lab at Stanford University in California used sophisticated genetic techniques on mice to turn on and off a class of neurons that includes chandelier

CELLULAR CULPRITS IN SCHIZOPHRENIA

According to one hypothesis, schizophrenia may trace back to defects in pyramidal and chandelier cells in the dorsolateral prefrontal cortex that lead to clinical symptoms only in adolescence.

REF. 1

Dorsolateral

Anterior cingulate gyrus

Ventromedial

Orbital

Cortical layers
1
2
3
4
5
6
White matter

Pyramidal neuron

Chandelier cell

cells — and found that this started and stopped much of the animals' organized PFC activity[6].

Lewis thinks that in people who later develop schizophrenia, chandelier cells fail at some crucial task of cultivating pyramidal cells during childhood or early adolescence. Together they do not generate the organized neural traffic required for building robust connections; and later, the weakness of both cell types leaves the PFC incapable of creating the vigorous, coordinated firing — including a type of electrical activity called gamma synchrony — that generates working memory. The result, subtle but increasingly apparent as synapses are pruned during adolescence, is a brain that can't consistently organize either its electrical activity or its thoughts: the shattered mind of schizophrenia.

Tom Insel, the director of the National Institute for Mental Health in Bethesda, Maryland, says that Lewis's model of this chain of dysfunction "provides something this field really needed: a framework for linking observations at the molecular, cellular and systems levels. We haven't had a story that crossed those levels of explanation before. And his story, whether it pans out in all its details or not, is invaluable for doing that."

While still thinking big, Lewis is filling in some of those details. Chandelier cells, for instance, signal using GABA, an inhibitory neurotransmitter. Lewis is trying to figure out whether errant GABA dynamics in schizophrenia are a cause of chandelier cell dysfunction or a compensation for it. He recently found that aiming an experimental drug at certain GABA receptors in chandelier and 'basket' cells (another kind of interneuron) boosted schizophrenia patients' gamma synchrony — and showed signs of improving their working memory. He's now trying to decide whether to aim another drug at those receptors or search for a different lever to pull — something that would reveal more cleanly the links among GABA, chandelier cells and the dysfunctions of schizophrenia. "You constantly have to balance this kind of close work with the big picture," he says.

For Lewis, the balancing act means tracking and responding to other lines of research, such as the work of clinical psychiatrist Anissa Abi-Dargham at Columbia University, New York. Abi-Dargham is using brain-imaging tools to explore whether flaws in evolutionarily older, subcortical areas that use the neurotransmitter dopamine are driving developmental problems in PFC circuits, or whether the problems in the PFC alter dopamine function. Lewis considers these connections between research programmes another sign that schizophrenia study has advanced in the last decade or so. "Used to be," he says, "we just got tired of a hypothesis or hit a dead end and went onto something else. Now we're actually integrating hypotheses or testing one against another.

"Or if you want to put it another way," he says, laughing, "we're getting a bit more synchrony in these findings. The communication between the blind men is improving. The elephant is starting to come together."

RUNNING AHEAD

The focus on adolescent brain development that has been so valuable in research generates controversy among clinicians, however. Particularly contentious is the idea of clustering schizophrenia's early whisperings into a diagnosable 'prodrome' period during adolescence. (The term comes from a Latin word meaning 'running ahead'.) The disease is typically diagnosed in young adulthood (see 'Accent on youth').

The North American Prodrome Longitudinal Study, or NAPLS, is an eight-centre project formed in 2003 that has been testing ways to reliably diagnose people in such a prodrome stage and treat them with psychotherapy, cognitive training, family therapy or drugs in the hope of forestalling worsening problems. NAPLS's main diagnostic tool is a questionnaire it calls the Structured Interview for Prodromal Syndrome. It scores symptoms such as fragmented or unusual thoughts; family histories of psychosis; social or school troubles; and paranoia or other peculiarities of emotion, behaviour or thinking.

In 2008, the group reported on 291 adolescents and young adults that the questionnaire identified as being at 'very high risk' of developing schizophrenia or other psychotic disorders[7]. Within 2.5 years of screening, 35% suffered psychotic episodes. The NAPLS researchers say that a set of prediction algorithms derived from those results afterwards, and then run again on the same data, raises the screen's predictive accuracy to almost 80% — comparable, they say, to risk predictions for medical problems such as heart disease. The team says that this and other "strong evidence for the prodromal risk syndrome … raises the question of its evaluation for inclusion in *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)" or *DSM* — a move that would lead to wide use of such screens and interventions.

This gravely concerns some clinicians, who point to the unavoidable false positives. "You will inevitably tell people who are not really at risk of schizophrenia that they are 'at very high risk'," says Til Wykes, a King's College London clinical psychologist who specializes in mental health (see also page 165). The impact could go far beyond inappropriate use of antipsychotic drugs, she says. It could negatively affect how families, friends and the broader community treat that person, as well as their self-conception. "The anxiety this produces may even generate just the thing you're trying to protect against. And you're doing this to people who are what — 15, 16? This is a huge intervention to take with someone who may not be destined for schizophrenia."

Others feel that the benefits of improved treatment outweigh the risks of overdiagnosis. "I frankly don't understand this concern about diagnosing a prodromal period, and I find this concern about overtreatment misplaced," says William McFarlane, a psychiatrist who runs a prodromal diagnosis programme at the Maine Medical Center in Portland. McFarlane argues that inclusion in the *DSM* would bring consistency of diagnosis and treatment to more people.
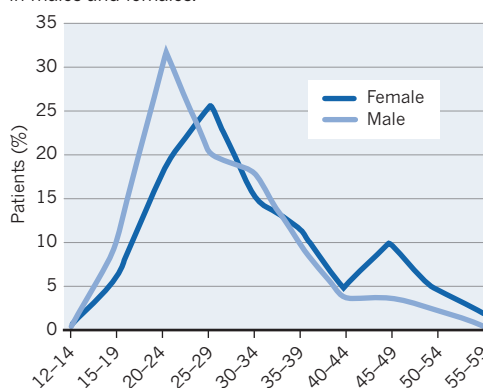
Many of the scientists untangling schizophrenia's complex developmental threads don't believe they yet have the tools to reliably discern a prodromal period — let alone treat it. "We have better ideas about what's going on than we did in the past," says Lewis, "but we do not yet have a target, or an intervention for that matter, that has a high enough likelihood of success."

Freedman thinks the state of knowledge requires caution and humility. "Schizophrenia research is full of people who are sure they know what they're doing, and only later do we understand that the whole paradigm was off. Then we look back in amazement at how wrong we had it. I like to think everyone in my generation would be well aware of this history, and be reluctant to say we're there." ∎

ACCENT ON YOUTH

Distribution of age at first admission for schizophrenia in males and females.

SOURCE: M. CANNON & P. JONES *J. NEUROL. NEUROSURG. PSYCHIATRY* **60,** 604–613 (1996).

David Dobbs is a freelance writer based in London. His website is http://daviddobbs.net.

1. Lewis, D. A. & González-Burgos, G. *Neuropsychopharmacology* **33,** 141–165 (2008).
2. Hashimoto, T. *et al. Biol. Psychiatry* **65,** 1015–1023 (2009).
3. Beneyto, M. & Lewis, D. A. *Int. J. Dev. Neurosci.* doi:10.1016/j.ijdevneu.2010.08.003 (2010).
4. Sweet, R. A., Fish, K. N. & Lewis, D. A. *Front. Hum. Neurosci.* **4,** 44 (2010).
5. Szabadics, J. *et al. Science* **311,** 233–235 (2006).
6. Sohal, V. S, Zhang, F., Yizhar, O. & Deisseroth, K. *Nature* **459,** 698–702 (2009).
7. Cannon, T. D. *et al. Arch.Gen. Psychiatry* **65,** 28–37 (2008).

THE DRUG DEADLOCK

The biology is too complicated.

Pharma companies are quitting.

Where are schizophrenia drugs going to come from?

BY ALISON ABBOTT

The shock wave hit when they broke the code. It was January 2005, nearly four years since the start of a clinical trial to definitively compare schizophrenia therapies. The US$43-million trial, involving nearly 1,500 patients at 57 clinical sites in the United States, was testing whether a raft of anti-psychotic drugs introduced in the 1990s — and hailed as transformational — was any better than a 50-year-old pill called perphenazine, one of a generation of drugs that left patients with horrible side effects. Until investigators unblinded the trial, codes concealed who was receiving which drug.

As it turned out, it didn't matter. The Clinical Anti-psychotic Trials of Intervention Effectiveness (CATIE) made it clear that the new therapies were barely different from the old[1]. They were just as good as perphenazine at controlling psychosis — hallucinations and delusions. But patients taking the new drugs remained confused, socially withdrawn and bereft of drive, just as they did on perphenazine. And the new antipsychotics were not even convincingly freer of side effects — overall, three-quarters of patients abandoned their drug during the 18-month treatment regime, regardless of which drug they took.

"That was frustrating and humbling for the research community," says Jeffrey Lieberman, a psychiatrist at Columbia University, New York, and the trial's principal investigator. "And it had a chilling effect on the pharmaceutical industry." Within a few years, under intensifying

pressure to rein in costs, several large companies, including London-headquartered AstraZeneca and GlaxoSmithKline, chose to pull out of psychiatric pharmacology altogether.

Chastened researchers also had to regroup. "It became a case of back to the drawing board," says Shitij Kapur, head of King's College London's section on schizophrenia, imaging and therapeutics. Scientists needed to learn much more about the disease's biology. They had to ensure that whatever they learned would be 'translated' more smoothly to the clinic, by way of better animal models, biomarkers and clinical trials. And they wanted to develop drugs to target not just psychosis, but also the 'negative symptoms' such as impaired cognition, blunted emotions and lack of initiative — the types of trait that render most people with schizophrenia incapable of holding down a job.

The scale of the work to be done was too daunting for individual labs. So in recent years, researchers working in the public and private sectors have decided to share more ideas and resources. Few expect a single molecule to do the entire job. "Fifteen years ago, we were naively optimistic," says Kapur. Now, there is still optimism — but with a hefty dose

SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

of pragmatism thrown in.

No one questions the transformational impact of the first antipsychotic drugs when they were introduced in the 1950s. Psychiatric hospitals could, for the first time, release large numbers of patients with schizophrenia who would otherwise have spent their lives incarcerated. The prototype, chlorpromazine, spawned a whole class of drug known as 'typical' anti-psychotics, including perphenazine.

MOVING TARGET
But the price of that freedom was high. Typical antipsychotics exert their effect by blocking the dopamine type 2 (D2) receptor, modifying dopamine neurotransmission. But these silenced receptors also provoked distressing side effects such as twitching and jerking, leading to the misconception that movement disorders and antipsychotic efficacy were inextricably linked.

So entrenched was this idea that when, in the 1960s, industrial pharmacologists discovered a promising antipsychotic drug candidate that did not disrupt movement in animal tests, they had difficulty persuading their sceptical managers to develop it. Sandoz, a company based in Basel, Switzerland, that is now part of Novartis, eventually introduced clozapine to the market in 1971. As well as acting on D2 receptors, it blocked the 5-HT$_{2A}$ receptor of the mood-modulating neurotransmitter serotonin, which seems to temper the movement side effects. The drug proved so much better that desperate

psychiatrists lobbied for its reinstatement after Sandoz withdrew it from the market in 1975 when its own rare side effect became apparent: a susceptibility to life-threatening infections.

The US Food and Drug Administration (FDA) relicensed clozapine in 1989 for treatment-resistant cases, in conjunction with regular blood tests. Within years, other drug companies had launched their own 'atypicals' (see 'Schizophrenia drug sales'), all with a clozapine-like pharmacology but intended to be safer. Yet some serious side effects, such as metabolic problems, emerged. By this point, clinicians were starting to question whether these new, and more expensive, drugs were any improvement on their predecessors. CATIE confirmed their worst fears.

A SELF-FULFILLING PROPHECY

A re-examination of the pharmacological profiles after the trial showed that although the atypical antipsychotics hit both D2 and 5-HT$_{2A}$ receptors, the D2 blockade seemed to be responsible for their clinical effects. That is not surprising in retrospect, given that the animal models used to test the drugs were all designed to pick up D2-receptor blockade, says Mark Tricklebank, a behavioural pharmacologist and director of the Lilly Centre for Cognitive Neuroscience in Windlesham, UK. "It was all very circular, a self-fulfilling prophecy," he says. "We'd been tuning the engine, when what we really needed was a new engine."

To break out of this vicious circle, scientists realized that they needed some fresh thinking in basic and translational science. Clinical trials had been mostly focused on treating psychosis, but there was increasing recognition that cognitive deficits — poor memory, inability to maintain attention and poor problem solving — were a fundamental aspect of the disease. In 2005, Steve Hyman, then director of the US National Institute of Mental Health in Bethesda, Maryland, launched Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS). The forum aimed to bring together academics, industry and the FDA to generate consensus about how best to design clinical trials to test drugs targeted at these cognitive deficits — and later extended to negative symptoms — through a battery of specially designed tests[2,3].

Guided by the MATRICS recommendations, several cognitive enhancers are already in early-phase clinical trial as potential add-ons to standard antipsychotic therapies. Also in clinical trial are several candidate drugs acting on receptors of the neurotransmitter glutamate — the only approach to have shifted focus away from dopamine. But the glutamate circuitry in the brain may prove hard to manipulate safely. And for the scientific community, the real challenge lies in understanding the system they are tinkering with. "We don't even understand schizophrenia at the biological level," says Thomas Laughren, the FDA's director of psychiatric drugs, voicing a frustration felt by many.

A European collaboration of researchers known as Novel Methods Leading to New Medications in Depression and Schizophrenia (NEWMEDS) is throwing every cutting-edge technology available at the problem in a very unusual public–private collaboration. Launched last year, the five-year, €20-million (US$28-million) effort funded by the European Commission includes seven academic partners, nine pharmaceutical companies (including AstraZeneca) and a few biotech companies. One of these is Icelandic genomics company deCODE genetics, which in 2008

SCHIZOPHRENIA DRUG SALES

The worldwide market for schizophrenia drugs has grown tenfold since the introduction of atypical antipsychotics such as Risperdal. Annual prescriptions have also risen dramatically.

SOURCE: IMS

identified in a large population study three variable genetic regions called copy number variations (CNVs), which, although very rare, confer a high risk of schizophrenia[4].

Scientists from industry say that it was initially hard to convince their companies of the value of sharing information in NEWMEDS. But they did — and to their evident glee, industrial pharmacologists can for the first time discuss openly, at least within the consortium, their individual approaches in psychiatric disease. Academic members, in turn, are gleeful about access to some of the industrial resources now on the table. They have pooled extensive information and material including data from many clinical trials in schizophrenia, sometimes with associated blood samples. "We now have the biggest database ever on this disease — more than 10,000 patients," says Tine Bryan Stensbøl, a pharmacologist at the Danish pharmaceutical company H. Lundbeck and the coordinator of NEWMEDS.

In a joint project led by geneticist Hreinn Stefansson of deCODE genetics, psychiatrist

Andreas Meyer-Lindenberg from the Central Institute of Mental Health in Mannheim, Germany, and neuroscientist Michael Brammer from London's Institute of Psychiatry analysed magnetic resonance imaging data from 500 people identified by deCODE genetics as having the high-risk CNVs, as well as 500 control subjects. They hope that the study will identify brain structures that are disrupted by abnormal genetic signatures and might eventually point to new therapeutic targets. "We wouldn't be doing this without the NEWMEDS initiative," says Meyer-Lindenberg.

An essential but admittedly less glamorous task for NEWMEDS is to determine the robustness of methods used to test drug candidates, particularly the animal and human tests of memory, attention and other aspects of cognition that are notorious for their sensitivity to tiny differences in environment. Unreliable tests may explain why drug candidates that look hopeful in animals fail in the clinic. Industrial and academic scientists are now using standardized protocols in their own labs, then comparing results and trying to understand why some may vary.

The consortium is also adopting rodent touch-screen technology, in which animals in behavioural tests tap a screen with their nose to get a reward for performing the experimental tasks, rather than press a lever or poke their noses into a hole. Being automated, it does not need constant observation. And crucially, results from such tests are potentially easier to translate into human psychology testing, which is increasingly based on touching screens.

The new concerted strategies could renew industry's optimism, even if there are few concrete signs of it just yet. With up to 1% of the world's population estimated to be affected by the disease, schizophrenia represents a huge potential market for any company that can find a new drug that genuinely improves any symptoms — particularly given that most patients develop the disease in their early twenties, and could be on daily therapy for the rest of their lives.
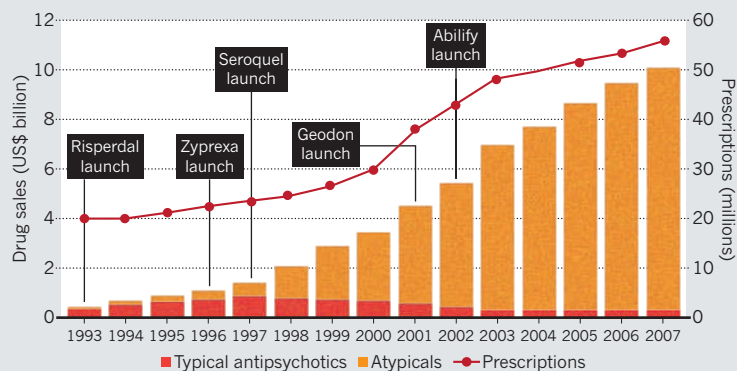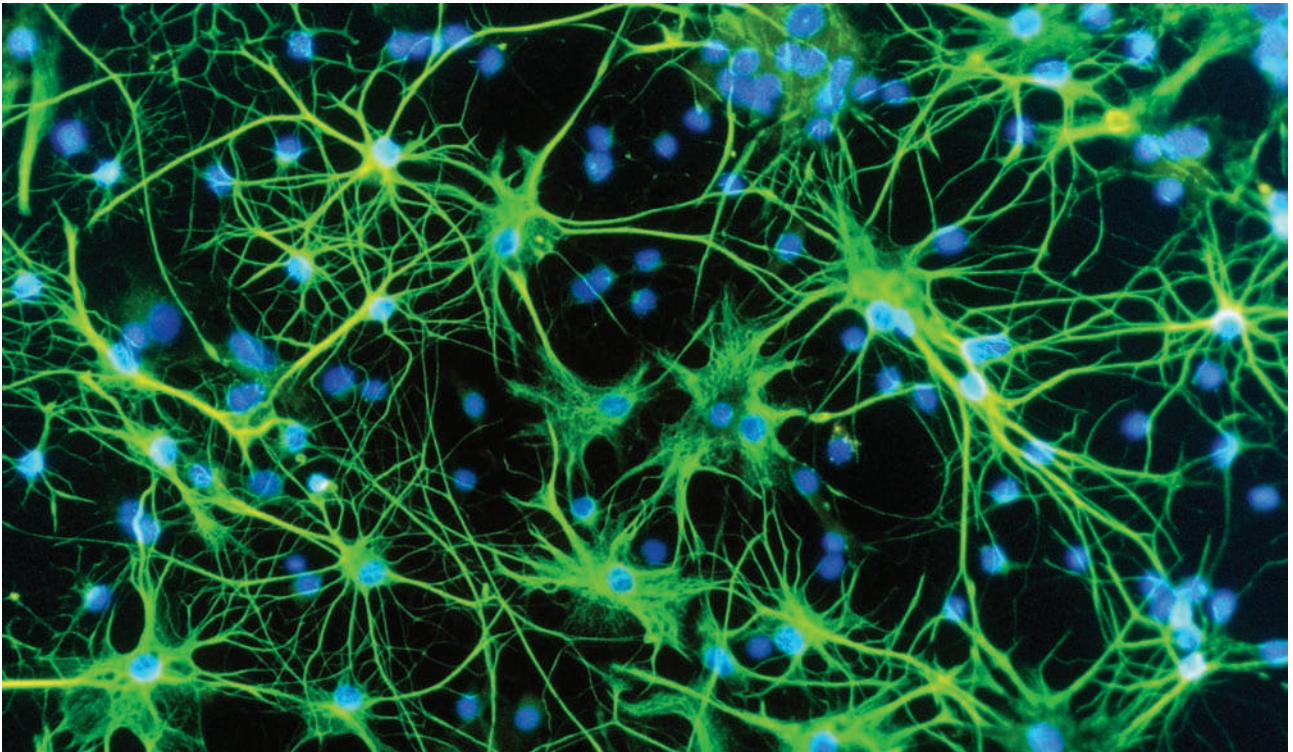
Maybe, after all, a shock wave was just what the field needed. ■

Alison Abbott is Nature's senior European correspondent.

1. Lieberman, J. A. et al. N. Engl. J. Med. **353,** 1209–1223 (2005).
2. Buchanan, R. W. et al. Schizophr. Bull. **31,** 5–19 (2005).
3. Kirkpatrick, B., Fenton, W. S., Carpenter, W. T. Jr & Marder, S. R. Schizophr. Bull. **32,** 214–219 (2006).
4. Steffanson, H. et al. Nature **455,** 232–236 (2008).

Astrocytes (stained green) form rich networks in the brain, which may increase the complexity of neural processing.

Settling the great glia debate

Do the billions of non-neuronal cells in the brain send messages of their own?

BY KERRI SMITH

Halfway through a satellite meeting at the Federation of European Neurosciences conference in Amsterdam in July, researcher Ken McCarthy takes the stage to give his presentation. He sports a black shirt and jeans, and his strong cheekbones, shock of white hair and tanned skin give him the look of a film star. But he doesn't have the confidence to match. "I find this a little bit daunting," he says, as he organizes his slides.

McCarthy, a geneticist at the University of North Carolina School of Medicine in Chapel Hill, is about to fan the flames of a debate about whether glia, the largest contingent of non-neuronal cells in the brain, are important in transmitting electrical messages. For many years, neurons were thought to be alone in executing this task, and glia were consigned to a supporting role — regulating a neuron's environment, helping it to grow, and even providing physical scaffolding (glia is Greek for 'glue').

In the past couple of decades, however, this picture has been changing. Some glia, known as astrocytes, have thousands of bushy tendrils that nestle close to the active junctions between neurons — the synapses (see 'Neural threesome'). Here they seem to listen in on neuronal activity and, in turn, to influence it. Studies show that chemical transmitters released by neurons cause an increase in the levels of calcium inside astrocytes, spurring them to release transmitters of their own. These can enhance or mute the signalling between neurons, or influence the strength of their connections over time. Moreover, astrocytes activated at one synapse might communicate with other synapses and astrocytes with which they make contact.

The consequences of this 'gliotransmission' could be profound. The human brain contains roughly equal numbers of glia and neurons (about 85 billion of each), and any given astrocyte can make as many as 30,000 connections with cells around it. If glia are involved in signalling, processing in the brain turns out to be an order of magnitude more complex than previously expected, says

Andrea Volterra, who studies astrocytes at the University of Lausanne in Switzerland. Neuroscientists, who have long focused on the neuron, he says, would "have to revise everything". In the past year or so, several papers have highlighted the urgency of this revision.

But the research that McCarthy is about to discuss could put a stop to the enthusiasm. "I'm presenting work from genetic studies that fly in the face of gliotransmission," he begins. Most studies so far have investigated astrocytes cultured in dishes, and bombarded them with calcium to elicit an effect. It has long been suggested, however, that these methods aren't specific enough to astrocytes, and might be affecting neurons as well. What is needed is a way to target astrocytes alone. So McCarthy has developed genetically engineered mice in which astrocytes can't signal normally. The mutations seem to have no effect on neuronal transmission in the brain.

His group's finding could come as a relief to some. Given the enormous neural complexity that gliotransmission would imply, "people don't want astrocytes to be involved", says Phil Haydon, a neuroscientist who studies glia at Tufts University in Boston, Massachusetts. But many, including most at the Amsterdam meeting, have built their careers on gliotransmission. In addition to their effects on the day-to-day functioning of the central nervous system, glia have opened new avenues of research into sleep, as well as psychiatric and neurological disease. Now, researchers are being forced to prove McCarthy wrong, or re-evaluate the fundamental precepts upon which the field was built. David Attwell, a neuroscientist at University College London, says that emotions in the community are running high. "If someone comes along and says that everything you've done is wrong, it's like you've wasted your life," he says. "It's become quite a polarized field, just in the last year or two."

The bomb

Gliotransmission has had plenty of support recently. In January this year, a group led by Dmitri Rusakov at University College London and Stéphane Oliet at the University of Bordeaux in France published results[1] suggesting that the chemical D-serine, released from astrocytes, activates a particular receptor — the NMDAR, or N-methyl-D-aspartate receptor — on the surface of neurons, influencing their behaviour. Communication through NMDARs is thought to be important in learning and memory, because it can help to enhance chatter between synapses and help memories to form.

And in September, a team led by Justin Lee at the Korea Institute of Science and Technology in Seoul found evidence[2] that astrocytes in the cerebellum release the neurotransmitter GABA (γ-aminobutyric acid). However, unlike the work of most groups, which suggests that astrocytes release chemicals packaged within tiny bubbles called vesicles, Lee's experiments imply that the cells are transmitting the chemicals directly through an ion channel in their membranes. When the researchers blocked the channel, called Bestrophin-1, GABA levels went down, suggesting that glia release GABA through this outlet.

But there have been concerns about experimental techniques. In these experiments, and many before them, investigators have tried to induce astrocyte signalling in petri dishes by pipetting calcium into individual cells and watching what happens. This approach is "like an atom bomb" going off in the cell, says McCarthy. Under natural conditions, the level of calcium would increase much more slowly, and McCarthy is concerned that pumping a cell full of calcium could make it behave strangely — perhaps forcing it to produce transmitters, or even preventing

it from releasing anything. McCarthy hoped to unpick the role of glia in the brain using his genetically engineered mice. "We thought this would definitively show gliotransmission," he says.

Instead, it totally crushed expectations. The team, led by McCarthy's postdoc, Cendra Agulhon, studied two mouse lines: one in which calcium signalling in astrocytes had been given a boost, and another in which it had been completely obliterated. But neither change made any difference to how nearby neurons were going about their business. McCarthy and his group were forced to conclude that astrocytes couldn't possibly be releasing chemicals to signal to neurons. They published their results in *Science* in March this year[3].

At the meeting in Amsterdam, McCarthy discusses his group's ongoing search for a behavioural effect that they could attribute to their wonky astrocytes. Again, no dice. "I would love to be able to show you that, but I can't," McCarthy says, amid murmurs of disbelief from the audience.

But some researchers have problems with his model. "The way he does it bothers me," says Richard Robitaille, a glia biologist at the University of Montreal in Canada. Measuring the effect of astrocytes will require more subtle experimental approaches, he says. Haydon agrees that it would be better to use more sensitive methods. He pictures astrocyte responses to stimuli as a narrow bell curve: if you blast them with something, they do nothing; if you don't stimulate them at all, they do nothing. But 'talk' to them using a low level of physiologically relevant stimulation and they should talk back. In McCarthy's mice, says Haydon, taking out the activity of all astrocytes probably changes so much — during brain development, and throughout life — that it is impossible to tell what the normal job of glia might be and how the brain compensates for their absence. The mice are thus subject to the same 'atom bomb' criticism levelled at cell-culture studies.

Part of the problem could be cultural. "The entire field has been trained in neurocentric labs, and everybody has so far believed that astrocytes work like neurons," says Maiken Nedergaard, a glial biologist at the University of Rochester in New York. "But astrocytes function totally differently. They use a different language. They use a different way of getting input and output." They may also work on a totally different timescale from neurons, says Rusakov. Their responses, he says, can be three orders of magnitude slower. As a consequence, common techniques for measuring neuronal responses won't work on astrocytes. Methods for imaging calcium in cells aren't good at measuring slow fluctuations or increases in the outer reaches of astrocytes, partly because calcium dyes simply don't penetrate there.

At his lab at the University of California, Los Angeles, Baljit Khakh has been developing a technique that can detect calcium in the previously inaccessible branches of astrocytes[4]. His team took an existing protein that is known to sense calcium, and modified it so that it could be targeted to cell membranes, where researchers suspect much of the calcium signalling might be going on.
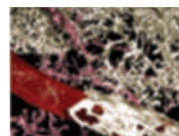
His work shows that a calcium increase in the main body of the astrocyte didn't necessarily cause a corresponding rise in calcium at the borders. If glia are indeed releasing transmitters, methods such as this should allow researchers to examine exactly where they are releasing them from.
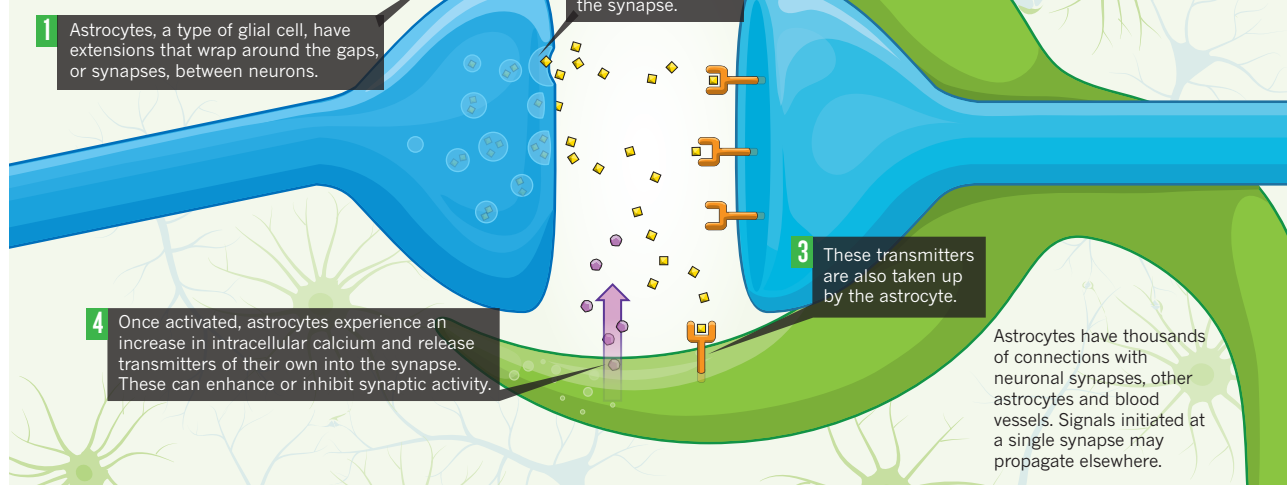
Robitaille aims to find out whether astrocytes can detect very low levels of synaptic transmission — a neuron passing just

THIS APPROACH IS "LIKE AN ATOM BOMB" GOING OFF IN THE CELL

NEURAL THREESOME

Several decades of study have focused on working out what is happening at the tripartite synapse.

1 Astrocytes, a type of glial cell, have extensions that wrap around the gaps, or synapses, between neurons.

2 One neuron signals to another by releasing neurotransmitters into the synapse.

3 These transmitters are also taken up by the astrocyte.

4 Once activated, astrocytes experience an increase in intracellular calcium and release transmitters of their own into the synapse. These can enhance or inhibit synaptic activity.

Astrocytes have thousands of connections with neuronal synapses, other astrocytes and blood vessels. Signals initiated at a single synapse may propagate elsewhere.

one electrical signal, or action potential, to the next. He and a colleague, Aude Panatier, used a high-speed imaging technique called line-scan imaging to measure calcium in astrocyte branches, and activated a small number of neurons using a very weak electrical pulse. Their results, not yet published, suggest that astrocytes can indeed pick up these low levels of activity and, what's more, they can regulate transmission themselves by releasing chemicals such as the energy-transfer molecules adenosine and ATP. Robitaille and his team have come down in favour of gliotransmission. "Based on our data, we have to sit on one side of the fence," he says.

Studies in different areas of the brain could also be contributing to the dramatically varying results. "Our knowledge about astrocytes is so poor that people tend to generalize findings in different circuits, different brain areas," says Volterra. "That also should be looked at." Even the time of day at which brains are prepared for experiments could be affecting measurements, says Haydon, because certain chemicals wax and wane over time. "I have a student go in and cut slices every four hours," he says at the conference, to a ripple of sympathetic chuckling from the audience.

The killer experiment
Methods aside, there doesn't seem to be a consensus as to what experiment would resolve this dilemma once and for all. "There's no clear simple experiment, otherwise I'd do it," says Attwell. "So would most of the others."

Most researchers in the field say that standardizing their methods and the ways in which they interpret their results could help. But the bigger personalities don't shy away from making ambitious statements. "Lots of neuron-centric people would look at our *Science* paper and say, 'That's the killer experiment,'" says McCarthy after his talk.

Later the same day, Haydon is gearing up to present some new results from his experiments on transgenic mice. When does he think the argument will be resolved? "Today," he says, "at 4.15."

Haydon's team works with mice with a different deficit to those that McCarthy is studying, although the two men — despite obvious differences of opinion — do share mouse resources and expertise. Haydon's mice have higher than normal levels of an enzyme called IP$_3$ phosphatase, which acts to prevent astrocytes releasing calcium. But the impairment in the mice that Haydon uses is more precise than that in McCarthy's animals, in terms of both timing and location. The calcium-blocking enzyme is expressed only after the mice are weaned, and only in a brain region called the hippocampus.

Haydon shows the assembled audience evidence that calcium signalling decreases in mice that have had their astrocytes disabled in this way, and that synaptic transmission in the hippocampus is also affected — implying that astrocytes help synaptic transmission to take place, under the right conditions. The work suggests that with heavy stimulation, synaptic transmission isn't changed; likewise, with a very weak stimulus, the astrocytes make no contribution. But there is a 'window of opportunity', at a medium level of stimulation, in which the effects are visible.

This is music to the ears of many glia enthusiasts. Even McCarthy is pleased to hear about positive results. "I would love to see gliotransmission," he says. But the evidence both for and against it still needs to be verified. With different groups approaching the question in different ways, and with the best-practice methods and techniques yet to be standardized and agreed on, it will take time before researchers can replicate each other's results and bring some clarity to the field.

McCarthy for one will keep searching for a behavioural effect in his mice with altered astrocytes, but is wary of letting what he wants to see taint his results. "If you get a vested interest in the answer, then you're in trouble," he says. Nonetheless, "If we do see it, we're going to be shouting it from the rooftops." ∎

Kerri Smith *is editor of the* Nature Podcast, *writing for* Nature *from London.*

1. Henneberger, C., Papouin, T., Oliet, S. H. R. & Rusakov, D. A. *Nature* **463,** 232–236 (2010).
2. Lee, S. *et al. Science* doi:10.1126/science.1184334 (2010).
3. Agulhon, C., Fiacco, T. A. & McCarthy, K. D. *Science* **327,** 1250–1254 (2010).
4. Shigetomi, E., Kracun, S., Sofroniew, M. V. & Khakh, B. S. *Nature Neurosci.* **13,** 759–766 (2010).

COMMENT

"DON'T JUDGE ME BY MY MENTAL HEALTH PROBLEMS"

LETS MAKE A DIFFERENCE

Short–lived campaigns are not enough

The stigma of mental illness will be reduced only if region–specific awareness initiatives become a permanent fixture of health and social services, argues **Norman Sartorius**.

While running the World Health Organization's mental-health programme from 1967 to 1993, I realized that the biggest barrier to progress was the negative attitude held by the public and decision-makers towards the mentally ill. On being elected president of the World Psychiatric Association in 1993, I therefore decided to make reducing the stigma of mental illness the focus of a major programme that would mark my time in office. Perhaps the most important outcome of that project — called Open the Doors — was the flaws it revealed in many of the assumptions on which national and regional anti-stigma campaigns are based.

Most people believe that those with mental illnesses are incapable of holding down jobs or maintaining relationships and, at worst, are dangerous, irrational and incurable. We know this from a wealth of anecdotal evidence and questionnaires conducted by mental-health experts[1,2]. This is despite studies showing that people suffering from depression, schizophrenia and other mental diseases perform well in jobs and in other social contexts, and are no more violent than the general population[3].

As well as making it difficult for patients

to find housing, partners, jobs and friends, this negative perception contributes to inadequate medical treatment. In a recent survey of people with schizophrenia, for example, 73% of more than 700 participants reported the need to conceal their diagnosis[4]. If people miss out on treatment because they are frightened about how friends and family will behave towards them when their diagnosis becomes known, the illness is likely to worsen. Untreated, patients stand a greater chance of getting embroiled in conflicts and being forced to enter a mental institution, which in turn contributes to the idea that mentally ill people are aggressive and irrational.

These pernicious misconceptions must be swept away. But short-lived anti-stigma ▶
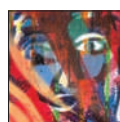
SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

▶ campaigns are not the best way to improve the situation.

In the past 20 years or so, governments, non-governmental organizations, charities and various institutions such as the World Federation for Mental Health have launched campaigns to combat the stigma of mental illness. 'Time to Change', for instance — run since 2009 by three British mental-health charities, Mind, Rethink and Mental Health Media, in collaboration with the Institute of Psychiatry in London — is the United Kingdom's largest effort yet to tackle the problem.

Many of these national or regional anti-stigma campaigns have followed a similar format. Messages to try to counteract people's prejudices are broadcast on the Internet, radio and television, often by celebrities. In general, the content of these messages — for example, that anyone can develop a mental disorder, that sufferers are not dangerous, and that mental illnesses are treatable — is recommended by psychiatrists, psychologists and other mental-health experts. In some cases, campaigns have included mental-health specialists giving lectures at schools or to the general public, for example about the implications and causes of mental illness. With some exceptions, such as in New Zealand and Canada, campaigns have been of limited duration, lasting from a few days to a year.

Such efforts have raised public awareness of the problem[5]. But there's no evidence to suggest that programmes lasting less than a year have had a serious, lasting effect on stigma.

REGIONAL DIFFERENCES

The World Psychiatric Association's Open the Doors programme began in 1996 and continued for more than a decade in most places. Participating countries included eight European countries plus Brazil, Canada, Chile, Egypt, India, Japan, Morocco, Turkey, Uganda and the United States[6]. Action groups of 6–8 people, including mental-health-care specialists, researchers, patients and their relatives, established what needed to be done to reduce the stigma of mental illness in their region and how to roll out events and implement changes in their area. Each group could draw on the resources and expertise of a larger advisory group of 30–40 people, including business people, journalists, ministries of social welfare, employees of social services and local celebrities.

By establishing a network of institutions and individuals committed to eroding stigma in their own and other countries, as well as a range of local changes to procedures and laws, Open the Doors reduced the degree to which people with mental illness are discriminated against and rejected by society all over the world[6]. What's more, despite the diversity of countries involved, several key lessons have emerged from the programme about how best to tackle the problem.

Celebrities, such as British actor and comedian Stephen Fry, have taken part in anti-stigma campaigns.

First, survey results should be used with caution. They should not be the only factor shaping a programme against stigma, nor provide the main way to measure effectiveness.

Assessments of general attitude — for example, using questionnaires that ask people whether they would employ someone suffering from depression or object to a mental institution being built in their neighbourhood — can provide valuable metrics. (A UK survey carried out by the Institute of Psychiatry every few years from 1993 to 2003 showed that people's attitudes towards the mentally ill had deteriorated over this period[7].) But a better way to establish which issues most need addressing in any given region is to ask people with mental illnesses and their families what is most burdensome for them.

In the Canadian arm of Open the Doors, for instance, people with mental-health illnesses and their families reported in focus groups that, for them, discrimination and insensitive behaviour shown by staff in health facilities was a far bigger problem than negative attitudes among the general population. So the Canadian action group prioritized working with health personnel, first in Calgary and then in Alberta. It held discussion groups to educate staff about ways in which staff behaviour might be improved — for example, by avoiding terms that people with mental illness can find insulting, such as 'schizophrenic'. The success of their approach led to Canada's hospital accreditation organization issuing national guidelines for general health facilities about how to treat those with mental illness[6].

Reporting in a questionnaire that you

"Education on the causes of mental disorders may even increase stigma."

would employ someone with a mental illness is not the same as employing them. As well as survey results, indicators of a campaign's success should include changes to employment legislation that diminish discrimination, changes to the proportion of the national budget devoted to mental illness, or reports from patients that they can more easily find jobs or housing than they could before the campaign began.

A second lesson that emerged from Open the Doors is that providing people with information, for example about the underlying cause of mental disorders and how they might change behaviour, may even increase stigma.

In Egypt, for instance, Open the Doors health workers tried to explain to Bedouins in the Sinai region that an illness needing treatment can cause certain types of behaviour, such as a peculiar way of speaking or reports of inexplicable voices and visions. Far from making them more tolerant, this new knowledge prompted relatives and others to start avoiding those individuals whom they now saw as sick. Bedouins suffering from a mental disorder received more help from their families and communities when the belief that some enemy had inflicted them with an evil spirit was allowed to persist. Similar cases of education leading to a subsequent increase or at least a failure to reduce stigmatization have been observed in Spain, Germany and South Africa[8].

In short, education and the use of surveys to track people's attitudes can be valuable components of anti-stigma work, but only as part of a slew of measures and efforts to improve the lives of patients and their families.

Perhaps the most striking lesson to emerge from Open the Doors is that the stigma of mental illness can't be held in check by bursts of effort, no matter how well implemented. Many of the people involved in national

T. WHITBY/GETTY IMAGES

and regional anti-stigma projects lasting a year or less have told me they felt let down at the end of the campaign — whether they are project organizers, volunteers, mental-health specialists or patients. No follow-up studies have been carried out to monitor the lasting effect of such short-lived campaigns, but people with mental illnesses especially report that after having their hopes raised, nothing had changed a year on.

To address the stigma of mental health in a meaningful way, strategies known to be effective should become a routine part of everyday services. These include educating health-care personnel, mediating face-to-face contact between the general public and people who have experienced mental illness, or persuading journalists to avoid certain terms when describing events relating to those with mental illness. Permanent networks of business people, journalists, social workers, mental-health experts, patients and volunteers need to be established. Also, goals should be tailored to local circumstances, by building and sustaining trusting relationships between everyone involved in the various anti-stigma efforts and patients and their families.

This is beginning to happen in some countries, including Germany, New Zealand, Brazil, Japan and the United Kingdom. For example, in Germany some of the education, media awareness and stigma research programmes started during Open the Doors are being continued on a long-term basis[9].

The overall lesson from Open the Doors is that stigma should be tackled in a fundamentally different way from most of the efforts carried out so far. Involving patients and their carers in the planning and evaluation of projects, for example, will result in a change in how programmes are constructed, funded and assessed. Likewise, including efforts to reduce stigma as a routine part of mental-health services will require a change in the organization and functioning of such institutions.

Stamping out stigma by altering the paradigms that have been the basis of most anti-stigma efforts will be difficult and costly. But doing so is crucial not only to improve the funding of mental-health programmes, the treatment of people with mental illness and their integration into society, but also to make our societies more civilized — a goal that concerns everyone. ∎

Norman Sartorius *is president of the Association for the Improvement of Mental Health Programmes at 14 Chemin Colladon, 1209 Geneva, Switzerland.*
e-mail: sartorius@normansartorius.com

1. Taskin, E. O. *et al. Soc. Psych. Psych. Epid.* **38,** 586–592 (2003).
2. Corrigan, P. *et al. Schizophrenia Bull.* **28,** 293–309 (2002).
3. Hopper, K. *et al.* (eds) *Recovery from Schizophrenia: An International Perspective* (Oxford Univ. Press, 2007).
4. Thornicroft, G., Brohan, E., Rose D., Sartorius, N. & Leese, M. for the Indigo study group. *Lancet* **373,** 408–415 (2009).
5. Vaughn, G. in *Mental Health Promotion: Case Studies from Countries* (eds Saxena, S. & Garrison, P.) 62–66 (World Health Organization, 2004).
6. Sartorius, N. & Schulze, H. *Reducing the Stigma of Mental Illness* (Cambridge Univ. Press, 2005).
7. Mehta, N., Kassam, A., Leese, M., Butler, G., & Thornicroft, G. *Br. J. Psychiat.* **194,** 278–284 (2009).
8. Angermeyer, M. C., Holzinger, A. & Matschinger, H. *Eur. Psychiat.* **24,** 225–232 (2009).
9. Baumann, A. E. *et al.* in *Understanding the Stigma of Mental Illness* (eds Arboleda-Flórez, J. & Sartorius, N.) 49–68 (Wiley, 2008).

Cognitive remediation therapy needs funding

More rigorous studies should be done on the effects of a therapy that seems to improve the everyday functioning of people with schizophrenia, says **Til Wykes**.

Cognitive remediation therapy seems to enhance the attention span, memory retention and problem-solving skills of those with schizophrenia through the performance of simple tasks. It has few side effects and is relatively cheap. A three-month course costs about US$700; three months' supply of the schizophrenia drug clozapine costs nearly $900. And there is growing evidence that the therapy could be at least as valuable as the drugs currently used to treat delusions and hallucinations, in helping people with schizophrenia lead fuller lives.

Yet investment in research has been lacking. In 2009, the US National Institutes of Health (NIH) spent nearly $350 million on schizophrenia research, but dedicated only $4.1 million of this to studies of cognitive remediation therapy.

To win the support of funders and turn remediation therapy into a serious treatment option, the community of researchers investigating it needs to develop a more rigorous approach. It is time to develop models of

Total National Institutes of Health spending on schizophrenia research in 2009

US$350 MILLION

Of which spent on cognitive remediation therapy
$4.1 MILLION

SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

how the training works, analyse how basic cognitive improvements help patients, and discover how best to implement the therapy and to whom it should be offered.

CLINICAL TRIALS
Medications for schizophrenia can reduce the degree to which patients experience delusions and hallucinations, but they are blunt instruments. Several medications recommended in national treatment guidelines, such as the UK National Institute for Health and Clinical Excellence, are associated with side effects such as weight gain or dribbling, and even worsened cognitive problems[1]. People's ability to lead a normal life is less affected by the degree to which they experience delusions and hallucinations following medication treatment than by problems with basic neuropsychological processes[2].

Various researchers worldwide began to develop cognitive remediation therapy in the late 1990s. The treatment involves mental-health professionals (social workers

or psychologists) coaching patients on tasks that boost memory, attention and other neurocognitive abilities. For instance, a patient might be asked to remember a set of numbers or pictures, and report them back in a certain order. They may then be shown how to improve their performance using a particular strategy, for example by splitting long lines of numbers into clusters of four.

More than 100 reports worldwide of controlled clinical trials involving some 2,000 people with schizophrenia now suggest that this type of training can improve the cognitive skills most closely tied to the ability to do things such as hold a job or develop relationships. Yet these studies are difficult to compare, and the resulting data tricky to pool. For instance, they use 99 different cognitive measures — among them, 'digit span' (the number of digits a patient can remember in a certain time) and 'processing speed' (how fast a person can manipulate and respond to information). Also, few models have been developed to explain how a therapy brings about cognitive changes, or why these changes affect a person's ability to function in everyday life.

Without this understanding, it is not clear which components of the therapy are likely to help a patient most. According to a recent meta-analysis[3], patients show greater improvement across a range of tasks when the therapy involves giving them the chance to practise. Yet their ability to function in everyday life is more closely tied to whether they are taught a strategy.

Another problem is that very few studies explore the differing effects the therapy has on different individuals. In one of the first studies of cognitive remediation for people with schizophrenia, I asked two male patients to say as many words beginning with the same letter as they could in a minute. Both participants improved their score, but one chose to call out words faster and made more mistakes in the process, whereas the other opted to make his choices more slowly and carefully. These differences in approach were associated with different activation patterns in brain scans[4].

By measuring whether people's performance improves, not whether it improves in the same way, we may be failing to tailor specific types of treatment to individuals — and so underestimating the overall effectiveness of cognitive remediation therapy.

SCIENTIFIC FOOTING

The research community must now take three key steps to put the study of remediation therapy on a more scientific footing, some of which are beginning to be taken by certain groups[5]. First, researchers must agree to use no more than ten standardized measures

of performance. A group of mental-health specialists in the United States recently agreed on a list of those aspects such as memory, attention and learning likely to be most important in judging the effects of cognitive remediation therapy. This list — the MATRICS cognitive domains[6] — offers a useful starting point.

Second, to look for individual effects, researchers should stratify the participants of studies into classes — according to age for instance, or ways of learning. To obtain the larger data sets needed to do this,

J. BURTON

researchers studying cognitive remediation therapy should agree to consign their data to a new collaborative database. (Currently most researchers place their data in the public domain, but there is no agreement over which variables to collect or how to merge different data sets.) This could operate in much the same way as the NIH-funded Psychiatric Genomewide Association Study Consortium — a collaborative effort to link biomarkers to certain mental diseases.

Third, researchers should investigate how cognitive changes are brought about and the effect they have on a person's life, in addition to looking at how much cognitive improvement occurs. Meta-analyses have shown that even small cognitive improvements can influence a person's ability to perform day-to-day skills. If these small improvements reflect changes in neural processing, it may be possible to identify the latter using brain-imaging techniques[7,8]. Also, certain properties of the therapy, such as how transferable a patient's acquisition of skills is to new settings — known as metacognition — may enable the transfer of basic cognitive improvements into better performance in everyday life. Training people with schizophrenia to decipher bus timetables or maps, for example, may boost the effects of treatment in the real world.

If these steps are taken, I estimate that, worldwide, a $20-million hike in funding over current levels will be all that is needed to bring cognitive remediation therapy into mental-health services. This is 0.05% of the US National Institutes of Health's entire annual spending and only 6% of its current total expenditure on schizophrenia research.

Measures of functionality such as employment indicate that, in the past century, there has been no improvement to the lives of people with schizophrenia[9]. Cognitive remediation therapy might improve this outlook, but only if we research and implement it in a truly scientific manner. ∎

Til Wykes *is professor of clinical psychology and rehabilitation at the Institute of Psychiatry, King's College London.*
e-mail: til.wykes@kcl.ac.uk
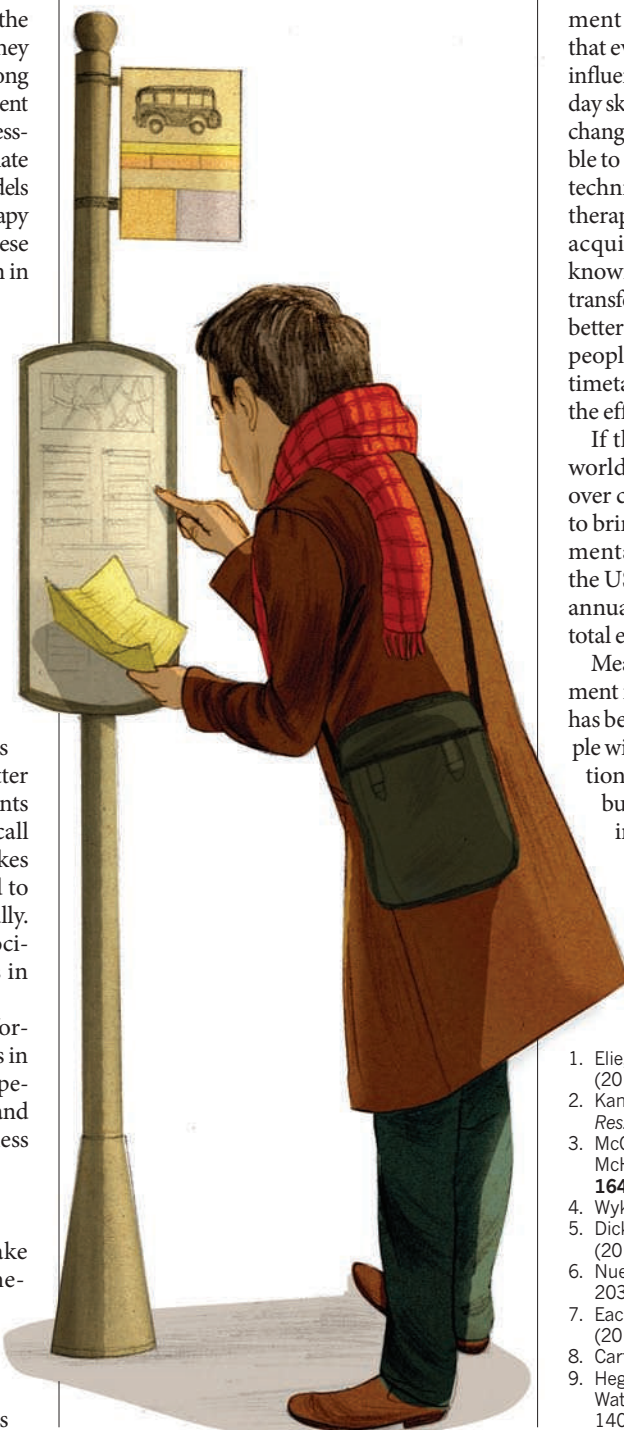
1. Elie, D. *et al. J. Psychopharmacol.* **24,** 1037–1044 (2010).
2. Kaneda, Y., Jayathilak, K. & Meltzer, H. Y. *Psychiat. Res.* **169,** 178–179 (2009).
3. McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J. & Mueser, K. T. *Am. J. Psychiat.* **164,** 1791–1802 (2007).
4. Wykes, T. *Schizophr. Res.* **34,** 77–86 (1998).
5. Dickinson, D. *et al. Am. J. Psychiat.* **167,** 170–180 (2010).
6. Nuechterlein, K. H. *et al. Am. J. Psychiat.* **165,** 203–213 (2008).
7. Eack, S. M. *et al. Arch. Gen. Psychiat.* **67,** 674–682 (2010).
8. Carter, C. S. *et al. Biol. Psychiat.* **64,** 4–10 (2008).
9. Hegarty, J. D., Baldessarini, R. J., Tohen, M., Waternaux, C. & Oepen, G. *Am. J. Psychiat.* **151,** 1409–1416 (1994).

Psychiatrists redefined many mental disorders after seeing the Second World War's effects on soldiers.

IN RETROSPECT
The five lives of the psychiatry manual

Roy Richard Grinker describes the military origins of the key reference work for diagnosing mental illness.

My grandfather, the pioneering psychiatrist Roy Grinker (1900–93), said that "to know schizophrenia is to know psychiatry". Indeed, psychiatry in the nineteenth and early twentieth centuries was largely an effort to understand psychosis, the majority of practitioners being employed in public mental-health institutions. The field broadened in the mid-twentieth century when psychiatrists became aware of a spectrum of mental disorders — not in hospitals, but in wars.

The Second World War led to major attempts to classify psychiatric conditions in the United States. Of the 11 million men and women who served in the US military between 1941 and 1945, 1 million were diagnosed with ill-defined 'neuropsychiatric' disorders. The inadequacy of clinical terminology at the time meant that a soldier with ordinary anxiety might be assigned a 'psychopathic personality' that clearly did

Diagnostic and Statistical Manual: Mental Disorders
AMERICAN PSYCHIATRIC ASSOCIATION
First published 1952

not apply. Yet these wartime syndromes were responsive to brief psychotherapy, often as simple as allowing the soldiers to talk about their experiences in a safe and restful environment. Psychiatrists' understanding of mental disorders and interventions expanded rapidly as a result.

In 1948, eager to advertise that psychiatry could treat more than psychoses, the Office of the Surgeon General published *Medical 203*, the US Army's classification manual for mental-health conditions. The volume was revised four years later by a committee of the American Psychiatric Association (APA), chaired by Captain George Raines of the US Navy, and published as the *Diagnostic and Statistical Manual: Mental Disorders* — now known as *DSM-I*. Released during the Korean War, *DSM-I* was a surprisingly

harmonious marriage of military experience and psychoanalytical theory.

Mental disorders were seen at the time as maladaptive neurotic reactions to the environment, and most included 'reaction' in their name. Disorders were classified in terms of the physical symptoms, the organ system involved (such as skin or cardiovascular) and whether the cause was known, secondary or unknown. A schizophrenic reaction, for example, was assumed to result from the individual's struggle to adapt to internal or external stressors.

OVERREACTION

The second edition, *DSM-II* (1968), retained the psychoanalytical focus on neurosis and adaptation. But it eliminated the use of 'reaction' in response to pressure put on psychiatrists to diagnose actual diseases, as other medical professionals did. 'Schizophrenic reaction' thus became 'schizophrenia'.

The 9 types of schizophrenia recorded in *DSM-I* were divided into 15 in *DSM-II* in order to capture a wide range of symptoms associated with the disorder — including some that appeared in other conditions, such as manic-depressive illness, depression and even bacterial infections. These 15 types included autism — then considered a feature of childhood-onset and paranoid schizophrenia rather than a distinct diagnosis — as well as subtypes that are no longer valid, such as 'latent type', which described early onset symptoms. However, the revision did little to standardize diagnosis. In a 1971 study, Robert E. Kendell and his colleagues showed that in cases in which the majority of US psychiatrists diagnosed a patient with schizophrenia, the majority of British psychiatrists diagnosed the same patient with manic-depressive illness.

A decade later, *DSM-III* (1980) revolutionized psychiatry, especially clinical trials and psychiatric epidemiology, by making it more evidence-based. Accepting that the physical origins of most psychiatric disorders were unknown (as they still are), the *DSM-III* authors eschewed psychoanalytical theory and hypothetical causes in order to establish diagnostic reliability and validity. Psychoanalysts were consequently outraged at what they considered to be a backwards step.

The standards in *DSM-III* were also aligned with the World Health Organization's International Classification of Diseases (ICD) manual, used in Europe. For scientists, these changes opened up new avenues for collaboration: researchers now shared the same language. For clinicians, the scientific criteria were a defence against the attacks on

SCHIZOPHRENIA
Search for origins and treatments
nature.com/schizophrenia

psychiatry as the subjective and dehumanizing profession represented in films such as *One Flew Over the Cuckoo's Nest* (1975). For patients, *DSM-III* promised more precision in diagnosis and treatment, especially for individuals who had rejoined their communities when many US public mental-health institutions were closed in the 1970s.

In the process of gathering evidence and aligning standards, subtypes lacking validity were collapsed. The schizophrenias dropped to five variants. And two main symptom domains for the disorder were established: the positive, which included hallucinations and delusions, and the negative, which noted impaired cognitive, emotional and social functions. The negative domain was especially beneficial because *DSM-III* could now capture individuals with schizophrenia who were not actively psychotic, or whose symptoms had changed over time.

Without the structure of underlying explanations, the manual became a list of symptoms for an expanding list of diseases, from a few dozen disorders in the first edition to well over 200. As psychologist Arthur Houts wrote, *DSM-III* showed how "a psychiatric nomenclature cut adrift from any theory became a nomenclature unconstrained".

For *DSM-IV* in 1994, the authors worked closely with authors of the ICD to make the two manuals congruent, clinically relevant, grounded in the most recent empirical research and more sensitive to how patients might interpret diagnosis. For example, in the updated *DSM-IV-TR* published in 2000, the writers eliminated language that might create stigma, replacing 'schizophrenic' — which implied that the disorder was an identity — with 'individual with schizophrenia'.

BACK TO THE FUTURE
Scientists now hope to integrate the classification of mental disorders with recent advances in genetics and neuroscience. Research suggests that many conditions, including schizophrenia, autism, bipolar disorder and depression, do not have distinct causes but arise during the brain's development, owing to shared genetic variations and relationships between behaviour and neural circuitry. The Research Domain Criteria (RDoC) project, launched by the US National Institute of Mental Health in 2009, is encouraging the study of mechanisms that are common to multiple disorders. Such approaches hark back to the causality on which *DSM-I* speculated.

Reconnecting our fragmented picture of mental illness is at the heart of *DSM-5* (Arabic numerals now replace the Roman), due to be released in 2013. Recognizing that narrow diagnostic categories do not help us to understand the way a person will develop over time, *DSM-5* will use symptom-severity scales instead of yes-or-no checklists to ▶

Books in brief
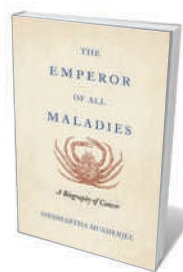
The Emperor of All Maladies: A Biography of Cancer
Siddhartha Mukherjee FOURTH ESTATE *592 pp. £25 (2010)*
The battle to cure cancer has been waged for thousands of years. From ancient Egyptian records of the illness to the latest research into how tumour cells multiply, physician and science writer Siddhartha Mukherjee relates our attempts to understand and control cancer. He asks why each new treatment for the disease — whether surgery, radiation or chemotherapy — ends up being applied at extremes that verge on the toxic. But he hopes that incremental knowledge will add up to transformative changes in treatment.
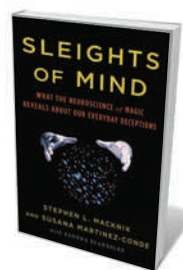
Sleights of Mind: What the Neuroscience of Magic Reveals about Our Everyday Deceptions
Stephen L. Macknik and Susana Martinez-Conde with Sandra Blakeslee HENRY HOLT *304 pp. $26 (2010)*
Good magicians know a lot about how the human brain works, and knowledge of magic can tell you a lot about the mind. The authors reveal the neuroscience behind the tricks of some of the world's great magicians. And using everyday examples of illusions, they explain how attention and awareness processes can be hacked. Practical applications range from the diagnosis of autism to marketing techniques and education.
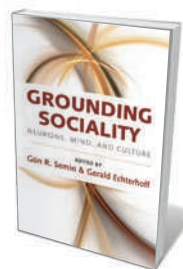
Grounding Sociality: Neurons, Mind, and Culture
Edited by Gün R. Semin and Gerald Echterhoff PSYCHOLOGY PRESS *288 pp. £45 (2010)*
Humans are social animals. The many aspects of how we interact on varying levels — from the neural to the behavioural and the cultural — are explored through contributions by experts from a variety of disciplines, edited by social psychologists Gün Semin and Gerald Echterhoff. The book asks how individuals take each other into account, coordinate their actions and share their inner thoughts, thus laying the foundations for an integrated view of sociality and its implications for the field of psychology.
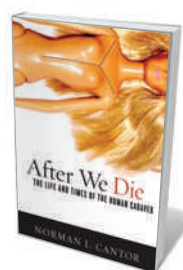
After We Die: The Life and Times of the Human Cadaver
Norman L. Cantor GEORGETOWN UNIVERSITY PRESS *284 pp. $26.95 (2010)*
Corpses have rights, argues legal scholar Normal Cantor. The wishes of the deceased regarding their mortal remains should be recognized and upheld, he reasons, suggesting that we should leave instructions for the disposal of our bodies. Examining the legal, ethical and moral aspects of death, he looks at the legacy of the corpse in medical education, science research and tissue transplantation. He also discusses novel endings for bodies, such as those that have been used in artistic displays or cryogenically frozen.
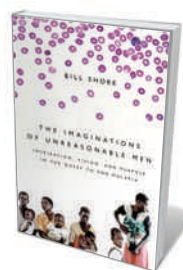
The Imaginations of Unreasonable Men: Inspiration, Vision, and Purpose in the Quest to End Malaria
Bill Shore PUBLICAFFAIRS *320 pp. $25.95 (2010)*
The scientists who search for a malaria vaccine inspire philanthropist and business leader Bill Shore's book. He uses the quest of these dedicated researchers as a springboard to muse on the lengths to which people will go to cure major world problems. Charting their efforts to overcome logistical and financial difficulties, and the disbelief of tropical-disease experts, he asks why and how these researchers persist in the face of adversity.

▶ better reflect the range and dynamics of patients' experiences. In another echo of *DSM-I*, special attention will be paid to context, in no small part caused by intense scientific and public interest in the mental-health problems associated with military personnel deployed in wars.

The number of possible diagnoses may contract in *DSM-5*. In the case of schizophrenia, the APA proposes removing all previous subtypes. The major debate is whether to add a category for psychosis risk syndrome, not unlike the 'latent type' within *DSM-II*. Some argue that because schizophrenia is a developmental disorder, with attenuated symptoms that can appear before psychotic episodes, a risk designation might aid early treatment and improve prognosis. Others note that most 'at-risk' individuals ultimately do not develop psychosis, and fear that the designation might turn normal human differences into pathologies or be motivated by pharmaceutical interests.

Each new edition of the DSM is considered a marker of progress, but we should be careful not to assume that psychiatric classification today is better than it was. Social scientists have widely criticized all releases of the DSM for making arbitrary distinctions between health and disease. They challenge the manual's power to dictate how the human mind is viewed across the public arena, in schools, hospitals and courts of law. Clinicians note that mental disorders are more heterogeneous than the DSM suggests, and question whether changes in classification have yielded better outcomes for patients.

Shifts in classification occur for many reasons — such as the influence of war, changes in the insurance industry or public attitudes towards mental illness — that have little to do with scientific progress and much to do with society and history. A disorder, even one with a clear cause or biomarker, is only a disorder when a society construes it as such. For example, Asperger's disorder is scheduled for elimination from the DSM, but this does not mean the category was wrong. It was useful when a non-stigmatizing term was needed for people with the disorder, but is becoming obsolete now that autism is accepted as a broad-spectrum illness without clear-cut subtypes. The possible collapse of schizophrenia classifications similarly reflects a more nuanced and connected picture of mental illness.

Nearly six decades after breaking ground with *DSM-I*, and three decades after *DSM-III* was radically reappraised, psychiatrists are braced for another diagnostic revolution. ∎

Roy Richard Grinker *is professor of anthropology and human sciences at George Washington University, Washington DC 20052, USA, and author of* Unstrange Minds. *e-mail: rgrink@gwu.edu*

Linked minds in a non-commercial setting can lead to ideas such as the Clifton Suspension Bridge, UK.

INNOVATION

Lighting the creative spark

Ingenuity combines individual skill with shared ideas, explains **Robert J. Sternberg**.

Two new books ask how we seed new ideas: Andrew Robinson examines the personal side of great breakthroughs in *Sudden Genius*, and Steven Johnson explores the collective basis of innovation in *Where Good Ideas Come From*. Both books highlight social and environmental factors in fostering creativity. But they tend to over-generalize, and fail to appreciate that many kinds of creative expression can emerge and be treated differently in society.

According to Robinson, most 'eureka' moments take a long time in coming. He describes psychologist John Hayes's finding that experts and geniuses must immerse themselves in a discipline for at least a decade before they have the knowledge and

Sudden Genius? The Gradual Path to Creative Breakthroughs
ANDREW ROBINSON
Oxford University Press: 2010. 352 pp. £18.99

Where Good Ideas Come from: The Natural History of Innovation
STEVEN JOHNSON
Allen Lane/Riverhead: 2010. 336 pp. £20/$26.95

experience needed to produce world-class work. Yet exceptional achievement clearly requires more than application: no amount of practice would allow most people to compose music like Mozart.

Asking to what extent genes determine creative success, Robinson observes that talent runs in families, suggesting that some aptitudes are inherited. Genius, however,

does not span generations, implying that it arises through a rare set of circumstances. He disputes the conclusion of polymath Francis Galton that such brilliance is inherited, as set out in his 1869 book *Hereditary Genius*, by arguing that Galton's study focused on people of talent more than genius. But he misses the point that Galton's early work was also flawed because it didn't account for environmental influences: family attitudes may have as big an effect as genes.

Robinson asserts that psychologists have produced no firm evidence that talent is largely innate. He leaves room for genetic factors but, for him, "determination, practice and coaching" are most important. There is clearly a correlation between practice and achievement. But a genetic predisposition might be more significant than Robinson supposes in helping an individual to reach the highest levels in a particular domain. Those lacking natural talent are more likely to drop out.

Talent, Robinson ultimately believes, results from an interaction of heredity and environment. On top of that, genius requires a combination of high motivation, personality factors such as openness to experience, immense amounts of learning, at least ten years immersed in one's discipline and an environment that

fits the potential genius like a glove. Many highly creative individuals, he also notes, lost a parent when they were young. This gives them an independent drive that might not have occurred otherwise.

Robinson leaves many questions up in the air, such as the relationship between intelligence and creativity, the role of the unconscious mind and links between genius and madness. There is evidence to suggest that intelligence (as defined narrowly by IQ) correlates with creativity up to an IQ of about 120, but has little effect beyond that. The unconscious does have a vital role in generating ideas, and a link has been demonstrated between bipolar disorder and certain forms of creativity, particularly in poetry.

Most provocatively, Robinson remarks that "talent appears to be on the increase, genius on the decrease". The reason for this shift, he suggests, is the increased specialization required for professional credentials, which makes the broad thinking that characterizes geniuses harder to develop. The problem is greater in the sciences than in the arts, owing to the higher technical skill base needed. And the astonishing amounts of complex knowledge that must be mastered today prevent most researchers from making deep connections between disciplines. Cross-disciplinary work, more and more, requires teams.

I believe that the culture of academia may also hamper genius. As philosopher of science Thomas Kuhn has pointed out, highly creative work that does not fit existing academic paradigms tends to be dismissed. Many great scientists have related how their most original ideas were repeatedly rejected by their peers.

Yet innovation can emerge from the collective, according to Johnson. In *Where Good Ideas Come From*, he argues that the world is better served by connecting ideas than by protecting them. He assigns creative ideas to four categories, according to whether they are the product of an individual mind or a network of minds, and introduced in a market or non-market setting. The most productive environment for the formation of ideas, he argues, is one that encourages networks of minds operating in a non-market setting. This category, he says, has yielded more great ideas than the others, including suspension bridges, anaesthesia, DNA forensics and the Krebs cycle for energy generation in aerobic respiration.

Universities, Johnson argues, have been the originators of many great concepts because they encourage the free interchange of ideas. To maximize creativity, you need both the availability of a network and the random collision of ideas within it, and universities offer both. He cites psychologist Kevin Dunbar's finding that few great scientific breakthroughs were the direct results of lab experiments; rather, those groundbreaking ideas arose during discussions afterwards.

Johnson, like Robinson, argues for the importance of the slow hunch, the perception that great ideas are gradual rather than fast in coming. Because they often build on platforms that others created — Isaac Newton's idea of standing on the "shoulders of giants" — those ideas are even slower in arriving than they seem. As well as serendipity, Johnson suggests that many good ideas emerge out of previous errors and 'exaptations': things that were created for one purpose but later turn out to be useful for another.

Although the two authors examine creativity from different perspectives — Robinson focuses on the individual, Johnson on the group — neither recognizes that there are different types of creativity and that the rules that apply to one type do not necessarily follow for others. For example, creative contributions that incrementally advance existing knowledge differ in their impact from those that redirect a field. The former are rewarded by a field's referees and editors; the latter, as Kuhn noted, may be accepted only grudgingly, if at all, as challenges to conventional wisdom.

> *"Flashes of brilliance are slow in coming. So too is their recognition."*

Thus it takes more than years of practice and the availability of a network to foster creativity. Inventive people tend to be crowd-defiers. Metaphorically, they buy low and sell high: generating an idea that, at the time, has little or no currency; convincing others of the value of the idea; and once their idea is accepted, moving on to the next unpopular idea. Creative people are thus intellectually combative.

A working environment that encourages creativity must tolerate and even encourage such contrariness. It must also recognize that the more creative an idea is, the harder it will be to sell. Reviewers of grant proposals and journal articles must recognize that highly creative research may be less developed than that which furthers established paradigms, and should make more allowances for originality.

Robinson and Johnson prompt us to think about where our ideas come from. If you haven't had a sudden revelation recently, don't worry — flashes of brilliance are slow in coming. So too is their recognition. ∎

Robert J. Sternberg *is provost and professor of psychology at Oklahoma State University, Stillwater, Oklahoma 74078, USA. His latest book is* Explorations in Giftedness, *co-authored with Linda Jarvin and Elena Grigorenko.*
e-mail: robert.sternberg@okstate.edu

Q&A Tracy Chevalier
On facts and fiction

Novelist Tracy Chevalier describes her experience of judging the entries in this year's Royal Society Prize for Science Books, and explains why she placed a nineteenth-century female fossil hunter at the centre of her last novel.

How did you find being on the judging panel?
It was daunting at first. I admire science, but feel inadequately prepared for it. As a student at Oberlin College in Ohio, I took science courses designed for non-scientists, but I haven't had any formal exposure since. When I read for leisure, I turn to novels. So imagine how it felt having to read 136 non-fiction books. Still, the older I get, the more I find myself reading non-fiction. At some point in your life you stop listening to pop music so much and start appreciating jazz.

How did you feel about the winning entry?
Nick Lane's *Life Ascending* [W. W. Norton, 2009] was beautifully written. As a novelist, I liked the overarching structure. It starts with the origins of life and ends with death. It follows a narrative that we are hard-wired to understand because it is our own life cycle. Along the way, each chapter is a self-contained building block, but it all fits together. There are also themes that keep reappearing throughout the story, all elegantly woven together.

What did you think of the other books?
On the whole, I was impressed with the quality of the writing but not with the production. Publishers seem to have a difficult time knowing how to sell science. Some of the back covers used language that was too user-friendly, calculated not to scare people off. At the other extreme were books that made no attempt whatsoever to be approachable.

Were any themes more popular than others?
There were loads of books about evolution with Darwin in the title, and lots about climate change. And quantum physics was big: we judges kept asking, why is a topic that is so incredibly difficult such a popular choice? I was overwhelmed by the largeness and smallness of the concepts the books covered, on time frames that are so far removed from the scale of our fourscore and ten years.

In your 2009 novel *Remarkable Creatures*, you fictionalized the life of Victorian fossil hunter Mary Anning. Why did you choose her?
I first came across Anning while visiting a small dinosaur museum in Dorchester, UK, where she's still seen as a local heroine. I was fascinated by her ability to make her mark in an arena dominated by middle-class men when she was a working-class woman. It was amazing that she could make important fossil discoveries that changed the landscape of natural science in the early nineteenth century without having been formally educated or trained. She received little recompense or credit for her findings.

Why write fiction about factual events?
I understand why historical fiction makes some people uneasy, and why they might prefer biography. In my novels I often point out what's true and what's not in an afterword. I try to be as accurate as I can. For example, I wanted Anning to have been present at the big Linnean Society meeting in London where her fossils were presented to the scientists. But she could not have been: we know she only went to London once, and that was later. So I decided to sneak her friend Elizabeth Philpot into that meeting instead. We have no idea whether or not Philpot was there in real life, so I felt I could make her do what Mary could not.

Does the theme of a woman getting an unfair deal resonate with you?
Anning once wrote that her brain had been "sucked dry" by men of science. Hers was a wonderful story. I like to write about people who are not normal, who are moving at the margins of society, who aren't happy with their lives, perhaps because they have no money or no power. That's why Anning was so compelling.

The shortlist contained no books by women writers. Does that bother you?
It was depressing. We judges didn't discriminate against female authors — only a few of the 136 books were written by women, so they never stood much of a chance. We need to encourage more women to write popular-science books. Given that many popular-science writers are also academics, perhaps women in those careers are spending their time publishing papers rather than writing books.

Do you think book awards are a good idea?
Prizes are crucial in giving books a boost. Hundreds of thousands of books are published each year, and we need something to help us distinguish what is really good — otherwise the choice is overwhelming. Prizes bring publicity and a prime position in bookshops. This is especially important for science books, which aren't always a natural draw to the public. They need that extra help. ∎

INTERVIEW BY JENNIFER ROHN

CORRESPONDENCE

Cities need plants and animals too

Your 'Science and the city' special issue (*Nature* **467**, issue 7318; 2010) overlooks the potential of urban areas as habitats for native species other than pigeons, rats and cockroaches.

Scientists and planners have traditionally treated human-dominated landscapes as incompatible with nature conservation. But the cost of city development need not include a complete loss of ecological function.

For instance, Norwegian-style green roofs, built with native plant species, would provide energy and allow reduction or collection of water run-off. Other ecological features, such as nesting cavities, could be incorporated into new buildings. Public spaces could host plants for native pollinators and provide habitat links within cities and with surrounding non-urban areas. Urban boundaries can offer resources to edge species that live in these transition areas, as well as to humans.

Patrick R. Huber, Steven E. Greco *University of California, USA, prhuber@ucdavis.edu*

Water: act now to restore river health

Urgent measures must be taken to ensure that it does not take a generation to muster the necessary political willpower to restore the health of rivers globally, as C. J. Vörösmarty and colleagues suggest (*Nature* **467**, 555–561; 2010). The livelihoods of 60 million people in the developing world depend on river fisheries, and millions more rely on them for food.

Lessons can be learned from Vietnam's positive actions, for example. Its Vu Gia–Thu Ban river basin is used for hydropower development, but rivers have dam-free stretches, designated after strategic environmental assessment with stakeholder participation (see go.nature.com/xnbbgr). Developing alternatives to large-scale mainstream dams — including river-bypass systems and micro-hydroelectric power installations for local use — will help to reduce trade-offs between water security and river biodiversity.

The pessimistic prediction of Vörösmarty *et al.* may well turn out to be correct for large river systems in Asia, Africa and Latin America, where weak governance of water allocation, dam construction and river management frequently overlooks the dependence of riparian communities on ecosystem health (P. Dugan *et al. Ambio* **39**, 344–348; 2010). Investment is needed now to build adaptive capacity and new livelihood opportunities for poor river communities.

We need much more investment of the kind made by the Consultative Group on International Agricultural Research. This group of publicly funded institutions is developing a programme to benefit poor people who depend on aquatic agricultural systems (see go.nature.com/g98kLv). This will focus initially on the basins of the Mekong, Zambezi and Ganges–Brahmaputra–Megna rivers.

Patrick Dugan, Edward H. Allison *The WorldFish Center, Penang, Malaysia, p.dugan@cgiar.org*

Water: biofuels sap supplies

The water footprint of the growing biofuel sector should be factored into discussions about water security (*Nature* **467**, 555–561; 2010).

The rapid expansion of biofuel crops can significantly affect regional hydrological patterns (B. G. Subhadra *Science* **329**, 1282–1283; 2010). In India, for example, jatropha plants — a biofuel feedstock with a large water footprint (W. Gerbens-Leenes *et al. Proc. Natl Acad. Sci. USA* **106**, 10219–10223; 2009) — are increasingly being cultivated in rural areas.

Add to this the rapidly depleting groundwater in northwest India (M. Rodell *et al. Nature* **460**, 999–1002; 2009) and the alarm sounds for prompt policy planning by the government to safeguard India's water resources.

Bobban Subhadra *University of New Mexico, USA, bsubhadra@salud.unm.edu*

Ireland should fund best research

The Irish government aims to develop our 'smart economy' by prioritizing funding for research that will lead to "innovation" (*Nature* **467**, 895; 2010). This will be achieved through another committee, the fourth in as many years (see go.nature.com/wepy1w).

The composition of this committee suggests that the choice of areas to be funded will probably be driven by commercial demands. That is unlikely to create new jobs, as the government claims.

Ireland should instead concentrate funding on research — any research — that is of consistently high quality (based on non-exchequer funding and prestigious publications) and dump the rest. Such a move would also attract high-quality researchers into the country.

More important than any ill-defined concept of innovation, it is crucial that high-calibre but inexperienced researchers get to work on problems with commercial potential, supported by consultants and professionals from industry. Top-class researchers need to be socialized, which is next to impossible in an academic setting.

The cuts to science in Ireland, combined with the government's handling of the remainder, risk turning the clock back 30 years.

Gareth Dyke *University College Dublin, Ireland, gareth.dyke@ucd.ie*

Concrete helix recalls smallpox win

The left-handed concrete DNA helix at the Australian National University's John Curtin School of Medical Research (JCSMR) in Canberra is symbolic of one of the world's greatest medical accomplishments (*Nature* **467**, 920; 2010).

A former JCSMR director, microbiologist Frank Fenner, chaired the World Health Organization Global Commission for the Certification of Smallpox Eradication. Smallpox virus was finally eradicated 30 years ago using a live vaccinia virus vaccine.

Crucial for transcription and for pathogenicity of poxviruses is Z-DNA's left-handed double-helical structure (A. Rich and S. Zhang *Nature Rev. Genet.* **4**, 566–572; 2003). Vaccinia virus becomes apathogenic if Z-DNA binding is disrupted (T. A. Brandt and B. L. Jacobs *J. Virol.* **75**, 850–856; 2001).

The Z-form of DNA on the outside of our JCSMR building is an inspiring reminder of the fundamental work done here — and that great discoveries can arise from researching unusual and anomalous phenomena.

Julio Licinio, Simon Easteal, Ma-Li Wong *The Australian National University, Canberra, Australia, julio.licinio@anu.edu.au*

John Huchra
(1948–2010)
Astronomer who mapped the structure of the Universe.

John Peter Huchra worked more hours in a day and observed more nights in a year than anyone else I know. In doing so, he helped establish our modern picture of the Universe as a frothy place, with sheets and filaments of galaxies enveloping giant empty voids. He also pioneered a technique to obtain more accurate measurements of the Hubble constant — a step that ultimately helped astronomers to pin down the age of the Universe as 14 billion years.

Huchra was born in 1948 to a train-conductor father and housewife mother in a poor neighbourhood of Jersey City, New Jersey. As he moved from success to success in academia, his humble beginnings remained a source of quiet pride. As an undergraduate, he helped to cover his expenses by unloading trailers. Even as a Harvard professor and member of the National Academy of Sciences, he said he continued to pay his union dues to the Teamsters so he could drive a truck if astrophysics didn't work out.

Huchra arrived at the California Institute of Technology in Pasadena as a PhD student in 1970 with the idea of becoming a theorist, but was quickly drawn to observational astronomy. He made observations using the most modern equipment and the most antique. He was the last person, for instance, to use the perilous Newtonian focus at the top of the Mount Wilson 100-inch (2.5-metre) telescope above Pasadena. This involved sitting in the dark on a platform perched at the top of the telescope, making observations while the enormous machine slowly pivoted to follow the target.

WORKING THE REDSHIFT

In the late 1970s, Huchra joined forces with Marc Aaronson at the University of Arizona and Jeremy Mould at the Kitt Peak National Observatory, both in Tucson. The team used a new method — infrared measurements of galaxy brightnesses — to calculate more precise distances between our Milky Way and other stellar systems. Legend has it that Huchra observed 130 nights in a single year on this project. This dwarfs the observing time allocated to a good-sized astronomy department and is a decade's worth of observing for many less-energetic astronomers.

The Hubble constant is the ratio of the velocity of galaxies (which are moving away from us as the Universe expands) to their distance from us. Huchra became a master at measuring that velocity, or redshift.

S. MITCHELL/HARVARD UNIV.

By 1984, the distances derived by his work with Aaronson, Mould and their associates produced evidence for a relatively high value of the Hubble constant: about 91 kilometres per second per megaparsec ($km\,s^{-1}\,Mpc^{-1}$).

A strong theoretical argument against such a high value for the constant came from the ages of the stars. If the Hubble constant were as high as Huchra and his colleagues claimed, the Universe would be younger than the oldest stars in our Galaxy. Huchra and his colleagues were not cowed by senior observers who advocated much lower values for the Hubble constant — or by theorists who believed that the Universe must be decelerating, which would make the age discrepancy even worse. They just reported what they found.

He and his colleagues improved their measurements of distances to nearby galaxies as technology permitted, and eventually the pieces came together. The modern value of $73\,km\,s^{-1}\,Mpc^{-1}$ emerged from studies using the Hubble Space Telescope in which Huchra participated. Together with the surprising finding that the expansion of the Universe is speeding up, the age of the Universe, and the ages of the stars in it, now fit into a plausible history of cosmic evolution that began with the Big Bang 14 billion years ago.

Huchra's other outstanding contribution was to generate three-dimensional maps of the distribution of galaxies using redshifts. He began this work as a postdoc in the 1970s and pursued it for decades.

In the 1970s, conventional wisdom among astronomers held that galaxies were either in clusters or evenly distributed in 'the field' between the clusters. But Huchra, Margaret Geller and their colleagues at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, discovered a different picture.

Earlier redshift surveys had indicated the possibility of large empty voids. As Huchra and his colleagues accumulated hundreds, then thousands, then tens of thousands of redshifts, their maps revealed that a foamy structure of voids, sheets and walls was a generic feature of the Universe. The team obtained this extraordinary picture through a confluence of insightful planning, new detector technology, creative analysis techniques and an ample dose of hard work at the telescope, much of which was supplied by Huchra.

John's energy and generosity extended far beyond his own research programme. As well as helping numerous PhD students to become successful scientists and people, he served on an astonishing number of panels and committees. He was president of the American Astronomical Society from 2008 to 2010. Most recently, he had a leading role in the National Research Council's Astro2010 Decadal Review — so was key to setting the agenda for the next ten years of astronomy research in the United States. In all of these activities, John was fair, thoughtful and realistic, seeking not the perfect outcome, but a good one.

> *"He was happiest in the observatory, with the controls of the telescope in his hands."*

John married late in life, and his joy with Rebecca Henderson and his pride in his son Harry made him a more complete person. Nobody worked harder at his craft, gave more of himself to his colleagues and students, and was less puffed-up by his considerable achievements than John. He was happiest in the observatory, with the controls of the telescope in his hands. On cloudy nights, he was unbeatable at 8-ball and pinochle, having misspent hours of his youth playing both. Counting cards, he said, is like counting galaxies. ∎

Robert Kirshner *is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. He was a fellow graduate student with John Huchra at the California Institute of Technology and a colleague at the Center for Astrophysics for the past 25 years.*
e-mail: kirshner@cfa.harvard.edu

NEWS & VIEWS

Recovery inhibitors under attack

Once a blood vessel supplying the brain has been blocked, the opportunity to prevent brain damage is fleeting. An alternative strategy might be to guide the damaged area onto the path to recovery. SEE LETTER P.305

KEVIN STALEY

The old saying that 'timing is everything' certainly applies to the treatment of strokes — the devastating injuries caused by obstructions to blood flow in the vessels that supply the brain. Advances in removing such obstructions have led to dramatic ameliorations in stroke damage. Unfortunately, however, restoring blood flow is beneficial only if achieved within the first few hours of obstruction: most patients still endure the full brunt of the injury. As a result, a rapidly expanding area of research is focused on enhancing the recovery from this damage, which is considerably less time-critical[1]. On page 305 of this issue, Clarkson and colleagues[2] report that recovery from a stroke is significantly improved when neuronal inhibition is reduced, starting three days after the injury.

Recovery of function after a stroke involves substantial alterations in the connectivity of the neurons that formerly interacted with the injured area[3]. Not surprisingly, many of these neurons are found in the brain regions immediately adjacent to the injury — the peri-infarct zone. A key regulator of the plasticity of synaptic connections is the inhibitory system subserved by the neurotransmitter GABA[4]. Clarkson et al. measured in vitro the currents mediated by GABA in mouse neurons of the peri-infarct zone, starting their measurements three days after a stroke. They found that the fast synaptic signalling mediated by GABA was normal, but that the background activity (referred to as tonic inhibition) was significantly increased in the peri-infarct zone.

Tonic inhibition operates through $GABA_A$ receptors, which are distributed between synapses to increase the permeability of the neurons' membrane to anions. This increase in the electrical leakiness of the neuronal membrane short-circuits the excitatory signals conducted from synaptic inputs to the cell body and to the axon hillock — the site at which action potentials are generated. The GABA that activates these extrasynaptic receptors arises from spillover at active synapses. GABA transporters normally return this GABA to both astrocytes and neurons (Fig. 1). But when neurons are depolarized, the transporters can function in reverse, causing extracellular GABA to increase.
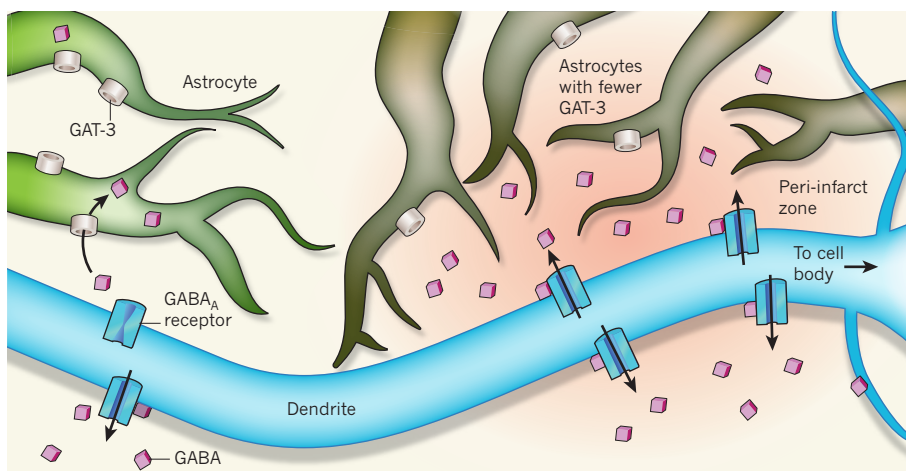
Figure 1 | The local consequences of stroke. Obstruction of a blood vessel (not shown) supplying the brain can cause a stroke, injuring the surrounding brain area — the peri-infarct zone. Clarkson et al.[2] find that, in the peri-infarct zone, loss of GAT-3, a transporter of the inhibitory neurotransmitter GABA, from astrocytes increases extracellular GABA levels and so activation of the $GABA_A$ receptor. This increases the shunting of excitatory currents in the dendrites of neurons, reducing activity-dependent neuronal plasticity, which is essential for post-stroke recovery.

GABA transporters might also function in reverse near the injured zone, as a secondary consequence of excessive activation of glutamate receptors or compromised energy production[5]. But when Clarkson et al.[2] blocked these transporters, they observed a robust increase in the GABA-mediated tonic conductance, suggesting that in the peri-infarct zone GABA transporters decrease — rather than increase — the extracellular GABA concentration.

There are several molecular species of GABA transporter. On sequential blocking of the transporters with selective pharmacological antagonists, the authors found that the GAT-1 transporter functions normally in the peri-infarct zone, whereas GABA uptake by the GAT-3 transporters is reduced (Fig. 1). They also report that a selective decrease in GAT-3 expression — and so not a change in the ionic conditions that drive uptake — is associated with changes in GABA uptake, increasing tonic inhibition in the peri-infarct zone.

Seizures, which can be caused by reduced inhibition, complicate 5% of the acute strokes that occur in the cortex region of the brain[6]. So an increase in tonic inhibition in the peri-infarct zone might be protective. But Clarkson et al. reasoned that such protection might come at the cost of reduced ability to alter synaptic connections[4] — a process essential for functional remapping of the cerebral cortex. Instead, reducing neural inhibition might enhance the processes of plasticity by which cortical regions are altered to improve muscle control. Indeed, the authors find that reducing tonic inhibition improves motor recovery after stroke, as evidenced by the number of missteps rats take while walking along a suspended wire grid.

Furthermore, Clarkson and co-workers present two independent lines of evidence to show that reducing tonic inhibition improves functional recovery after a stroke. In one set of experiments, the authors studied mice lacking two GABA-receptor subunits that are found primarily in extrasynaptic receptors and so mediate tonic inhibition. In another set, they used an antagonist that specifically targets these subunits to selectively reduce GABA-mediated tonic inhibition. In both cases, the outcome was improved post-stroke recovery of locomotor function.

These intriguing observations[2] open many research avenues. First and foremost is safety: because of the incidence of seizures after stroke[6], the risks of reducing GABA-mediated

tonic inhibition must be carefully assessed. The source of the increased tonic GABA conductance in the peri-infarct zone also warrants further investigation; pharmacological removal of GAT-3 in the healthy neocortex does not alter tonic inhibition[7].

Again, timing is everything: previous work[8] indicated that increased inhibition at the time of the stroke is beneficial. In addition, when Clarkson et al. reduced inhibition too soon after the stroke, they observed a detrimental effect — increase in stroke size. These timing constraints point to overlapping beneficial and detrimental GABA functions and need to be resolved.

Clarkson and colleagues' experimental reductions of GABA-mediated tonic inhibition affected the entire brain, leaving uncertainty as to whether the enhanced tonic inhibition in the peri-infarct zone was the main site of action. Strokes alter the activity of local networks in which the injured zone was involved[9]. Consequently, both local and general alterations in inhibition might help to reconstitute some of these network activities[10] and so improve functional recovery.

All of the benefits that Clarkson et al. report for reducing GABA-mediated inhibition had occurred by the time of their first assay of functional recovery — one week after the stroke. Thereafter, both treated and untreated animals recovered at the same rate. This result again raises the possibility that a reduction in GABA-mediated tonic inhibition improves the function of the damaged cortical networks, as opposed to enhancing the long-term recovery of those networks. The latter would increase the rate of improvement in gait, and would persist after the GABA blockers are removed. Of course, it could be that both effects (improvements in immediate function as well as recovery) contribute to the observed improvements in muscle control: Clarkson and co-workers show that discontinuation of GABA blockade removes about half of the improvement in recovery.

Reducing GABA-mediated inhibition enhances alertness. Although Clarkson et al. rule out an immediate performance enhancement by treating a subgroup of animals just before each test, stimulants are known to improve stroke recovery in rodents[11], if not humans. So future work should carefully control for effects of GABA manipulations on the level of consciousness.

Strategies to accelerate recovery from stroke not only offer a possible complement to the emergency rescue strategies, but are also much more feasible: they can be used at later times after a stroke. The present study[2] promises one such strategy, subject to further investigation. ∎

Kevin Staley *is at the Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA.*
e-mail: kstaley@partners.org

1. Hachinski, V. et al. Stroke **41,** 1084–1099 (2010).
2. Clarkson, A. N., Huang, B. S., MacIsaac, S. E., Mody, I. & Carmichael, S. T. Nature **468,** 305–309 (2010).
3. Cramer, S. C. Ann. Neurol. **63,** 272–287 (2008).
4. Martin, L. J. et al. J. Neurosci. **30,** 5269–5282 (2010).
5. Moskowitz, M. A., Lo, E. H. & Iadecola, C. Neuron **67,** 181–198 (2010).
6. Camilo, O. & Goldstein, L. B. Stroke **35,** 1769–1775 (2004).
7. Keros, S. & Hablitz, J. J. J. Neurophysiol. **94,** 2073–2085 (2005).
8. Green, A. R., Hainsworth, A. H. & Jackson, D. M. Neuropharmacology **39,** 1483–1494 (2000).
9. Paz, J. T. et al. J. Neurosci. **30,** 5465–5479 (2010).
10. Sanes, J. N. & Donoghue, J. P. Annu. Rev. Neurosci. **23,** 393–415 (2000).
11. Sprigg, N. & Bath, P. M. W. J. Neurol. Sci. **285,** 3–9 (2009).

ELECTRONICS

A diverse printed future

An approach that entails printing compound–semiconductor ribbons on a silicon substrate offers the means to build nanoscale transistors that can be switched on and off much more effectively than their bulk analogues. SEE LETTER P.286

JOHN A. ROGERS

For applications in electronics, silicon is often referred to as 'God's material'. Characteristics such as high natural abundance, and relative ease in crystal growth, purification and doping, combine with favourable electronic-transport properties to provide unmatched capabilities for commercial integrated circuits. As a result, silicon has held a dominant position in microelectronics since the early days of the industry, with compound semiconductors, used mainly in radio-frequency devices, consistently a distant second.

Sometime in the latter part of this decade, however, fundamental limitations on the switching speed and energy efficiency of silicon transistors may force a shift to a certain level of diversity in semiconductor materials[1,2]. One future approach might involve integrating non-silicon semiconductors onto silicon platforms, to yield heterogeneous systems that exploit different types of materials for different functions. On page 286 of this issue, Ko et al.[3] report an intriguing route to this goal, which is based on organized arrays of ribbons of indium arsenide (InAs) delivered to silicon wafers in a type of printing process[4]. Transistors built with such ribbons at nanoscale thicknesses exhibit impressive characteristics, suggesting their potential for enhancing the performance of next-generation silicon electronics.

Compound semiconductors such as InAs are attractive because their extremely high electron mobilities and conductivities lead to transistors that can be faster (up to twice as fast) and more power efficient (up to ten times) than silicon transistors with comparable dimensions[2]. Although poor hole mobilities (where a hole is a 'missing electron') and lack of high-quality, interfacial insulators will probably prevent their exclusive use in large-scale complementary logic circuits[1,2], these materials have potential as strategic additions to silicon-based technologies. The most widely explored means of exploiting compound semiconductors in this fashion involve specialized procedures for growing or bonding these materials on silicon wafers. Although certain research demonstrations are encouraging, such strategies have serious shortcomings, ranging from defects in the materials to challenges in manufacturability. Ko et al.[3] present an advanced procedure that avoids these limitations, and they demonstrate their ideas with InAs.

In the first step of the procedure, Ko and colleagues exploit optimized techniques to grow pristine, ultrathin films of InAs on gallium antimonide (GaSb) wafers coated with layers of aluminium gallium antimonide (AlGaSb). Next, the authors pattern the InAs films into narrow, nanoscale-thick strips that they release from the underlying substrate by selectively removing the AlGaSb with a chemical etchant. In a final step, they use a silicone rubber stamp to lift arrays of the nanoribbons from the substrate, and then to deliver them to the silicon dioxide (SiO₂) insulator surface of a silicon wafer, in a type of printing process[4] in which the InAs serves as the 'ink' (Fig. 1, overleaf). Because the procedure can be used with different types of material, the authors refer to the resulting structure as 'X' on insulator, or XOI, where X represents a semiconductor, by analogy to the widely used acronym SOI for silicon on SiO₂/Si substrates.

The printing process used by the authors[3] represents a recent and increasingly sophisticated method for transferring nanoscale ribbons, wires and sheets of semiconductors (such as silicon, gallium arsenide, gallium nitride and indium phosphide) from substrates on which they are formed to other surfaces, including those of silicon, glass, plastic and even paper and rubber[4,5]. Demonstrated applications of the process include electronics integrated with biological systems[6], hemispherical 'eyeball' and near-infrared imagers[7,8], flexible display and lighting devices[9]
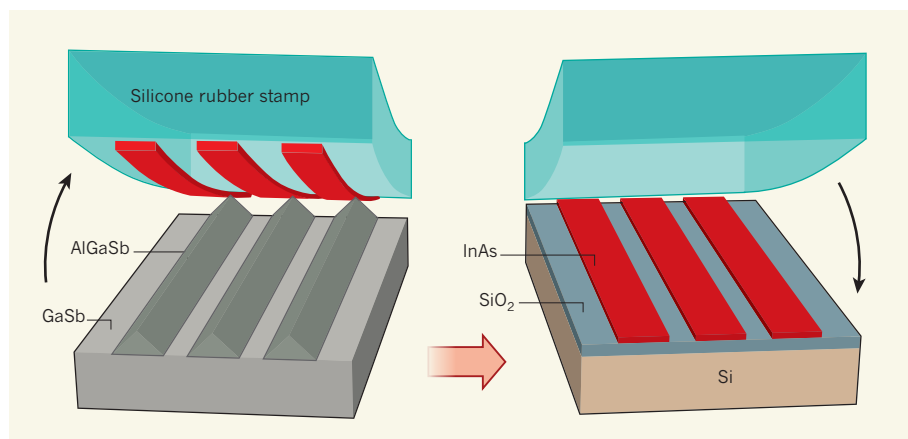
Figure 1 | Heterogeneous electronics by printing. Ko and colleagues' printing technique[3] for making heterogeneous structures for electronic applications uses a silicone rubber stamp to lift nanoscale-thickness indium arsenide (InAs) ribbons from a gallium antimonide (GaSb) wafer coated with a layer of aluminium gallium antimonide (AlGaSb). The nanoribbons are then delivered to a silicon dioxide/silicon (SiO$_2$/Si) substrate in a process in which the InAs acts as the ink.

and photovoltaic modules[8]. In many of these examples, viscoelastic effects[4] and/or specialized relief structures[10] on the stamps enable printing of pristine, unaltered material onto bare substrate surfaces, even without any separate adhesive layers. Yields approaching 99.99% are now possible with highly developed tools that also offer micrometre-scale precision in the positions of the printed parts, and throughputs corresponding to millions of printed structures per hour, or more[8].

These printing methods are presently in use for the pre-commercial manufacture of photovoltaic modules that incorporate sparse arrays of thin compound-semiconductor solar cells and micro-optics for focusing incident sunlight[11]. Although the same methods have been suggested for integrating compound semiconductors with silicon[4,5,12], Ko and colleagues[3] achieve by far the most impressive results in this context, accomplished by using semiconductor-material layers at exceptionally small thicknesses, down to just a few nanometres.

With remarkably clean, adhesiveless interfaces and high-quality, thermally grown oxides, these ultrathin semiconductor layers yield transistors that can be switched on and off much more effectively than their conventional, bulk counterparts. The authors[3] describe systematic experimental studies that capture the essential physics of operation of one such type of device, in which an interesting and gradual transition from three- to two-dimensional electronic transport occurs as the thickness decreases from 50 nm to less than 10 nm. Device simulations not only quantitatively capture these trends, but also explain related improvements in switching properties. This match between theory and experiment provides further evidence of the defect-free, predictable nature of the printed material stacks from which the devices are made.

The transistor's performance parameters are highly promising, with electron mobilities

that significantly exceed those of silicon transistors of similar design. The behaviour of the device at high switching speeds, however, must be evaluated to determine the potential for enhancing the performance of state-of-the-art silicon platforms. Exploring aspects of operation in this regime and demonstrating

interconnected operation with silicon transistors represent directions for future work. Research of this type is appealing because it advances knowledge in both science and engineering, in the context of potential solutions to problems of practical importance. The increasingly ubiquitous nature of electronics in modern society suggests that successful outcomes will have widespread, positive implications. ∎

John A. Rogers *is in the Department of Materials Science and Engineering, University of Illinois, Urbana, Illinois 61801, USA. e-mail: jrogers@illinois.edu*

1. Heyns, M. & Tsai, W. *MRS Bull.* **34,** 485–488 (2009).
2. Chau, R., Doyle, B., Datta, S., Kavalieros, J. & Zhang, K. *Nature Mater.* **6,** 810–812 (2007).
3. Ko, H. *et al. Nature* **468,** 286–289 (2010).
4. Meitl, M. *et al. Nature Mater.* **5,** 33–38 (2006).
5. Ahn, J.-H. *et al. Science* **314,** 1754–1757 (2006).
6. Viventi, J. *et al. Sci. Transl. Med.* **2,** 24ra22 (2010).
7. Ko, H. C. *et al. Nature* **454,** 748–753 (2008).
8. Yoon, J. *et al. Nature* **465,** 329–333 (2010).
9. Park, S.-I. *et al. Science* **325,** 977–981 (2009).
10. Kim, S. *et al. Proc. Natl Acad. Sci. USA* **107,** 17095–17100 (2010).
11. Burroughs, S. *et al. Proc. 6th Int. Conf. Concentrating Photovoltaic Systems,* April 2010, Freiburg, 163–166 (Am. Inst. Phys., 2010).
12. Benkendorfer, K., Menard, E. & Carr, J. *Compound Semiconductors* 16–18 (June 2007).

NEUROSCIENCE

The split view of motion

In both fruitflies and vertebrates, signals from photoreceptor cells are immediately split into two opposing channels in the downstream neurons. This might facilitate the computation of visual motion. SEE LETTER P.300

CHI-HON LEE

Nearly a century ago, the great Spanish neuroanatomist Santiago Ramón y Cajal compared[1] the vertebrate retina with the fly's compound eye and noted similarities in their neural circuits (Fig. 1). He redrew the cell bodies of the fly's monopolar cells, transforming them to vertebrate retinal bipolar neurons. Ultrastructural studies have since revealed that, indeed, both sets of neurons receive inputs from photoreceptor cells at structurally unique junctions called ribbon synapses in their first visual neuropiles, or neural switchboards — namely the fly's lamina and the retina's outer plexiform layer[2]. On page 300 of this issue, Joesch *et al.*[3] further extend the analogy, reporting that, like their vertebrate bipolar-neuron counterparts, fly monopolar cells split photoreceptor signals into ON and OFF channels to encode brightness increment and decrement, respectively.

Two main types of fly monopolar cell — L1 and L2 — receive a similar number of synaptic

inputs from the type of photoreceptors that mediate motion detection. Using genetic methods to manipulate the activity of specific neurons, behavioural studies[4,5] have suggested that L1 and L2 have overlapping but differentiable roles in detecting visual motion.

By recording electrical activity from downstream motion-sensitive neurons, Joesch *et al.* provide a physiological basis for the behavioural observations. They find that blocking L1 eliminates the response to a moving bright edge (ON), whereas blocking L2 abolishes responses to a moving dark edge (OFF). In a separate paper[6], the same group directly examines the activity of L2 neurons by calcium-imaging techniques and confirms that L2 encodes the OFF signals. Thus, as for vertebrate photoreceptors, the fly photoreceptor signal is split into ON and OFF channels at the first synapse.

Joesch and colleagues[3] further unexpectedly find that L1 and L2 are electrically coupled through gap junctions — specialized complexes that connect the cytoplasm of

two adjacent cells. This observation reconciles a conundrum of previous behavioural observations[4,5]: reconstituting photoreceptor input to either L1 or L2 is sufficient to generate a near-normal response to motion stimuli, whereas blocking the output of either L1 or L2 diminishes that response.

Why split photoreceptor signals into ON and OFF channels? It has been suggested[7] that this coding mechanism satisfies both information and metabolic constraints: to keep the downstream retinal ganglion neurons at a high level of basal neural activity costs energy; to inhibit them below a certain basal level conveys little information. But why do only those fly photoreceptors that mediate motion detection split their signals into ON/OFF channels in the lamina? The other photoreceptors, which mediate colour vision, do not. Joesch *et al.* argue that splitting facilitates the computation of visual motion, and to do so they invoke the original Hassenstein–Reichardt model for motion detection.

The Hassenstein–Reichardt model was developed[8] in the 1950s to explain behavioural experiments on the beetle *Chlorophanus*. Despite these humble origins, in its abstract form this model has arguably been the most influential theory of visual motion detection[9]. It explains how motion can be computed from local luminance changes — the fundamental form of visual motion. And it makes several counterintuitive predictions that match well with electrophysiological and behavioural data from many species, including humans.

The core computation of the Hassenstein–Reichardt model is based on a delay-and-compare mechanism: the signal from one photoreceptor is delayed and then compared, by multiplication, with the instantaneous signal derived from a neighbouring photoreceptor. Despite its simple construction, the multiplication stage is actually difficult to implement by known synaptic mechanisms, because this seemingly simple mathematical operation needs to be carried out in a way that respects the 'sign' of the signal — that is, multiplying two negative signals should generate a positive signal.

Joesch *et al.*[3] point out that the 'original' Hassenstein–Reichardt model (reported in German and consequently lost to most of the English-speaking world) offers a solution to this problem. In that model, photoreceptor signals are first split into ON and OFF channels, which carry positive and negative components, respectively. Four separate multiplications for the two neighbouring signals are then added or subtracted to correct the sign and generate a direction-specific motion signal. In a way, the sign of the signal is 'remembered' by the signal paths, and the outcome is identical to a sign-corrected multiplication. Electrical engineers in the 1960s devised a similar solution called the four-quadrant (Gilbert) multiplier, which is still used in analog circuits today.
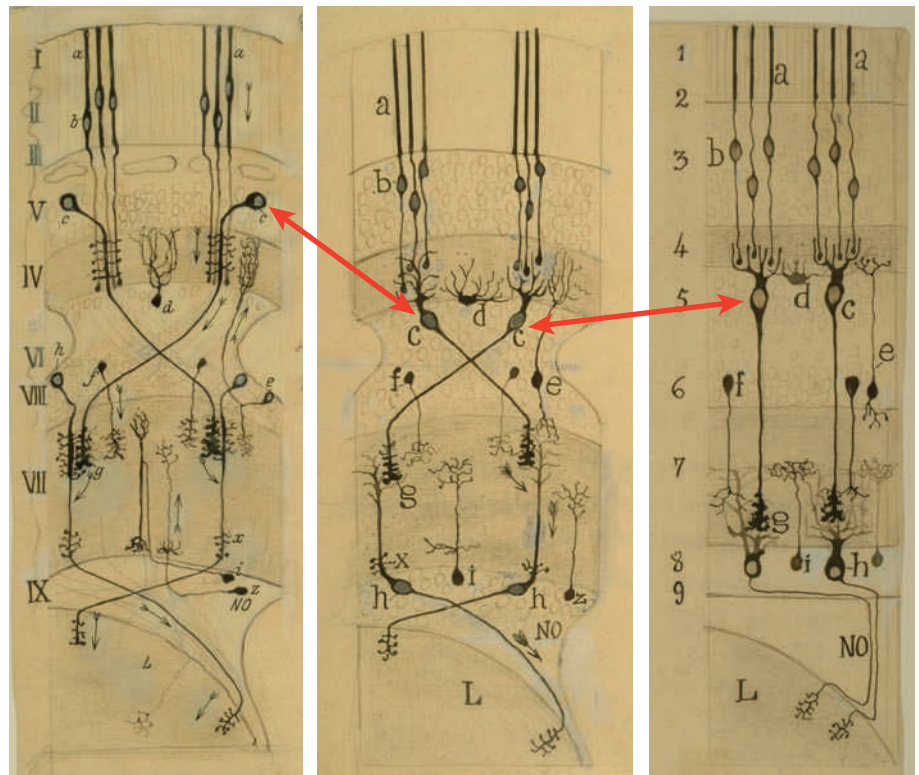
Figure 1 | Similarity between fly and vertebrate visual systems. Santiago Ramón y Cajal[1] compared the monopolar cells in the fly visual system (left panel) with the bipolar cells in the vertebrate retina (right panel). He redrew the cell bodies of the former to take on the bipolar form (middle panel) (as shown by the added red arrows). The 'mysterious' black arrows (left panel), which appear in many of Cajal's drawings, correctly indicate the flow of visual information, from the photoreceptors (a, b) to the monopolar cells (c) and to the downstream neurons (h).

An advantage of the Hassenstein–Reichardt model is its robustness: inactivating either the ON or OFF channel reduces but does not abolish the motion response, consistent with both behavioural and electrophysiological data. Over time, however, several motion-detection models have been proposed, including the energy model[10], which generates output identical to that of the Hassenstein–Reichardt model but has a very different internal structure. To determine which of these various models explains the behaviour of neurons mediating visual motion, direct access to the actual neural circuits is required.

As one cost of its abstract nature, the Hassenstein–Reichardt model is anatomically inexplicit and only hints at the actual neural implementation. To carry out four-quadrant multiplication, each elementary motion detector requires four separate pathways feeding ON and OFF channels into four multipliers, each of which converges on the wide-field motion-sensitive neurons from which Joesch *et al.*[3] recorded. Previous anatomical and electrophysiological studies[11,12] have revealed a number of candidate neurons for visual motion dectection. More are anticipated from ongoing ultrastructural projects to reconstruct these circuits *in toto*, especially those at the Janelia Farm campus of the Howard Hughes Medical Institute in Ashburn, Virginia. The recent convergence of anatomical, behavioural and electrophysiological investigations — all aided by powerful fly genetics — provides renewed hope that the neural mechanism of motion detection might finally be resolved in the near future. ∎

Chi-Hon Lee *is in the Program in Cellular Regulation and Metabolism, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA.*
e-mail: leechih@mail.nih.gov

1. Cajal, S. R. & Sanchez, D. *Trab. Lab. Invest. Biol.* **13,** 1–167 (1915).
2. Meinertzhagen, I. A. *Prog. Retinal Res.* **12,** 13–39 (1993).
3. Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. *Nature* **468,** 300–304 (2010).
4. Katsov, A. Y. & Clandinin, T. R. *Neuron* **59,** 322–335 (2008).
5. Rister, J. et al. *Neuron* **56,** 155–170 (2007).
6. Reiff, D. F., Plett, J., Mank, M., Griesbeck, O. & Borst, A. *Nature Neurosci.* **13,** 973–978 (2010).
7. Schiller, P. H. *Trends Neurosci.* **15,** 86–92 (1992).
8. Hassenstein, B. & Reichardt, W. *Z. Naturforsch.* **11b,** 513–524 (1956).
9. Borst, A. & Egelhaaf, M. *Trends Neurosci.* **12,** 297–306 (1989).
10. Adelson, E. H. & Bergen, J. R. *J. Opt. Soc. Am. A* **2,** 284–299 (1985).
11. Fischbach, K. F. & Dittrich, A. P. M. *Cell Tissue Res.* **258,** 441–475 (1989).
12. Douglass, J. K. & Strausfeld, N. J. *J. Neurosci.* **15,** 5596–5611 (1995).

ANIMAL BEHAVIOUR

How to confuse thirsty bats

Echolocating bats have a legendary ability to find prey in the dark — so you'd think they would be able to tell the difference between water and a sheet of metal. Not so, report Greif and Siemers in *Nature Communications*. They have found that bats identify any extended, echo-acoustically smooth surface as water, and will try to drink from it (S. Greif and B. M. Siemers *Nature Commun*. doi:10.1038/ncomms1110; 2010).

The way in which bats locate point objects has been studied extensively, but how they recognize extended objects, such as pools of water, isn't known. As pictured here, bats drink while on the wing. Greif and Siemers hypothesized that, when searching for a drink, the animals look for the echo-reflection signature of water surfaces — the only extended, acoustically smooth surfaces in a bat's environment.

When a bat sends an echolocation beam at a glancing angle to a water surface, most of the beam bounces off the surface away from

the animal, like light off a mirror. But a small part of the beam travels vertically down from its source, and is reflected right back to the bat. This reflection pattern could act as a flag for water.

To test this idea, the authors conducted experiments on 15 species of wild bat, placing them in a room that had two large plates on the floor. The plates were made of one of several materials: wood, metal or plastic. Each of the surfaces was either smooth or textured. The smooth surfaces reflect echolocation beams in the same way as water, and, sure enough, thirsty bats repeatedly tried to drink from these surfaces, but ignored the textured ones (see movie at http://go.nature.com/pnpal8). The authors thus concluded that bats use echolocation to recognize bodies of water.

When Greif and Siemers trialled juvenile bats that had had no previous contact with ponds, the animals also tried to drink from the smooth plates, thus revealing the
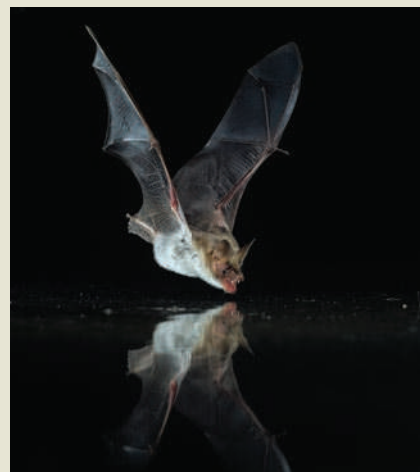
water-location mechanism to be innate. What's more, the authors found that echolocation overrides conflicting sensory stimuli such as vision, chemoreception and touch. For example, if a smooth surface was placed on a table, the bats tried to drink from it even if they had already flown under the table. The authors suggest that innate water recognition in bats could be used to study the neural basis of habitat recognition. **Stefano Tonzani**

D. NILL

FUNDAMENTAL CONSTANTS

Big *G* revisited

Measuring Newton's constant of gravitation is a difficult task, because gravity is the weakest of all the fundamental forces. An experiment involving two simple pendulums provides a seemingly accurate but surprising value.

RICHARD DAVIS

Newton's law of universal gravitation[1] is a pillar of classical physics. Here's a quick textbook example: the gravitational force between any two spherical objects is proportional to the product of their masses and inversely proportional to the square of the distance between their centres. If you know the value of each mass in kilograms and the distance between them in metres, the Newtonian constant of gravitation, *G* (aka big *G*), lets you calculate the gravitational force between the masses in units of … newtons! Big *G* is one of the fundamental constants of physics[2]. Its value, which is roughly 6.674×10^{-11} m^3 kg^{-1} s^{-2}, can be established only by measurement. However, experiments with the potential to yield a highly accurate value of *G* are notoriously challenging. In a beautifully written article in *Physical Review Letters*[3], Parks and Faller describe an experiment carried out at the JILA institute in Boulder, Colorado, that has allowed them to measure *G* with an uncertainty of 0.0021%, or 21 parts

per million (p.p.m.). This is among the smallest uncertainties ever achieved, but the derived value of *G* is a surprise.

The basic idea of Parks and Faller's experiment can be illustrated by a simple pendulum

(Fig. 1a). When a 'source mass' is brought near the pendulum's bob (the 'test mass'), the gravitational attraction between the two masses causes the bob to move a small distance, *z*, from its usual rest position. Of course, the design and analysis of the real experiment are much more sophisticated than this simple depiction. The authors' experiment has two pairs of tungsten source masses and two identical pendulums, the copper bobs of which are pulled in opposite directions, and a host of other clever features.

The distance each bob moves is small: *z* is of the order of 50 nanometres. Yet the authors show that such small displacements can be
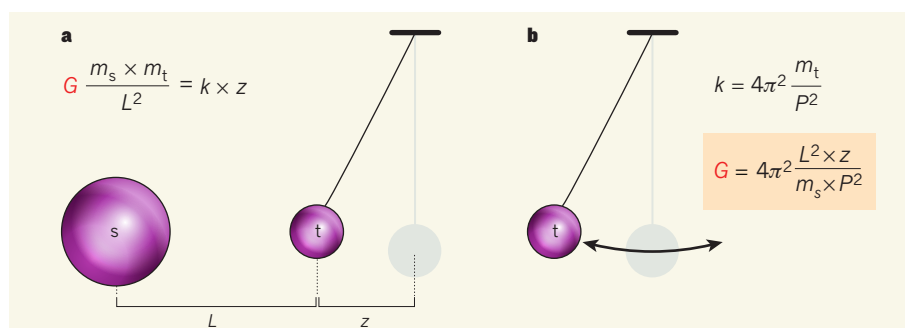
Figure 1 | The basic principle of Parks and Faller's experiment[3]. a, A spherical 'source mass' (m_s) is brought near a pendulum's spherical bob (the 'test mass', m_t) and causes the bob to move a small distance *z* from its usual resting position (grey). The gravitational force between the two masses (left side of equation), which depends on Newton's constant (*G*), can be obtained from a measurement of *z* provided that *k* is known (see **b**). **b**, The value of *k* is found by measuring the period (*P*) of the freely swinging pendulum. To compute the value of *G*, we need measurements of *L*, *z*, m_s and *P* (but not m_t). Parks and Faller's experiment was based on four cylindrical source masses of 100 kilograms each, two pendulums and many other refinements.

$$G \frac{m_s \times m_t}{L^2} = k \times z$$

$$k = 4\pi^2 \frac{m_t}{P^2}$$

$$G = 4\pi^2 \frac{L^2 \times z}{m_s \times P^2}$$

50 Years Ago

'Anatomical evidence for olfactory function in some species of birds' by Betsy Garrett Bang
It seems curious that the large olfactory organs of certain species have so often been pointed out by anatomists, yet most olfactory learning studies have been done on feebly equipped birds such as pigeons, and have tended to keep alive in the text-books the idea that the chemical sense in birds is minimal or lacking … In dissecting the nasal tissues of birds for work on natural defences against respiratory disease, I have been repeatedly impressed by the size of some of the olfactory conchae … Of the species thus far dissected, by far the largest and most heavily innervated organs have been seen in the turkey vulture, *Cathartes aura*, the Trinidad oilbird, *Steatornis caripensis*, and the Laysan and black-footed albatrosses, *Diomedea immutabilis* and *D. nigripes*, each of which types represents a separate order with quite dissimilar feeding and nesting habits … There is no question of degenerate or indifferent function of the olfactory organs of these birds.
From *Nature* 12 November 1960

100 Years Ago

'The Cocos-Keeling Atoll' — During a very short visit to these islands some years ago I was taken across the lagoon in a light canoe, and when wading to land, about a quarter of a mile distant, over the rough surface of fresh coral branches, I suddenly crashed downwards for about 2 feet into a mass of rotten coral which spread over an irregular area some 20 or 30 yards across. I did not investigate this further, as a shark's fin appeared above the water off shore, but Mr. Ross informed me that a good deal of the coral in the lagoon had been "killed" at various times by sulphurous exhalations from below.
From *Nature* 10 November 1910
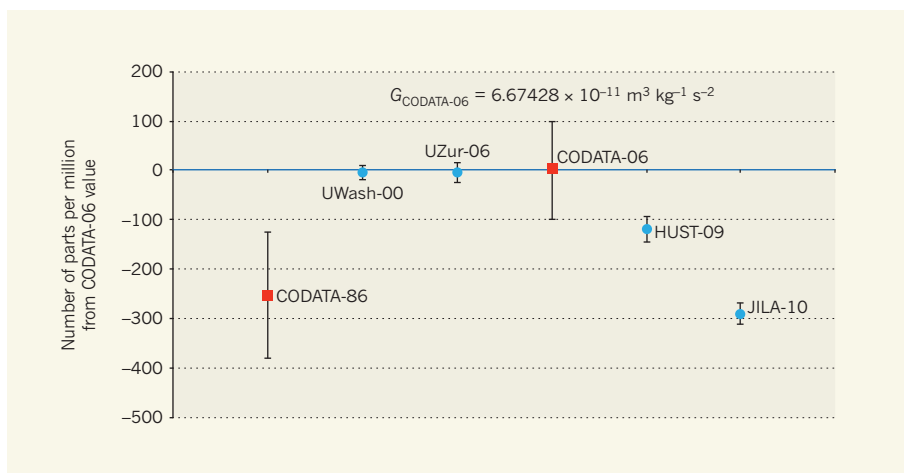
Figure 2 | Parks and Faller's estimate of big *G* in context. Parks and Faller find[3] a value for *G* (JILA-10) in agreement with the estimate of the CODATA-86 Task Group on Fundamental Constants but in disagreement with the CODATA-06 value[4]; from time to time, CODATA reviews and combines results from various experiments. Two of the measurements with the smallest uncertainty, UWash-00 (ref. 6) and UZur-06 (ref. 7), have been taken into account in determination of the CODATA-06 value. HUST-09 represents the culmination of an experiment that appeared in ref. 5 as HUST-99. All error bars denote 68% confidence levels.

measured with high accuracy. To measure the analogue of *z* in their experiment, Parks and Faller attached mirrors to the bobs and used modern optical techniques. Because precise optical measurements are impossible if the pendulums are swinging, the researchers installed powerful, permanent magnets beneath each bob so that a phenomenon called eddy-current damping would keep them still without affecting the values of *z*. But the magnets did create some small, subtle problems, which had to be identified and solved.

As a final step, the authors removed the magnets and source masses so that each pendulum could swing freely. This allowed the researchers to measure the period of each pendulum — the time it takes for the bob to complete one full swing — and, in turn, to derive the value of *G* from the measured distance (corresponding to *z* in Fig. 1b).

Here's the surprise: Parks and Faller's result[3] does not agree with the previous best estimate[4] of *G*, which was provided by the CODATA Task Group on Fundamental Constants (Fig. 2). CODATA regularly publishes an in-depth review of relevant experiments, followed by a list of recommended values and uncertainties for the fundamental constants of physics, including *G*. The last such publication, CODATA-06 (ref. 4), considered all results that were available until the start of 2007. (Earlier reports were dated 2002, 1998, 1986 and so on.)

To put the authors' work in context, it is helpful to know a bit of the recent history of big-*G* measurements. Typically, the set of credible *G* results available to CODATA is not consistent[5] for reasons that are seldom clear. Nevertheless, CODATA must produce its recommendation. In 1995, a new and highly discrepant experimental result led CODATA to increase the uncertainty assigned to *G* from

130 p.p.m. (1986) to a whopping 1,500 p.p.m. (1998), although it decided that the recommended value of *G* should not be changed. This unsatisfactory situation was a call to action, eventually leading to many new experimental results. By 2005, CODATA had sufficient reasons to exclude the discrepant 1995 value from further consideration[5].

The CODATA-06 recommended value for *G*, and the four experimental results that have the smallest estimated uncertainties ever reported, are shown in Figure 2. Interestingly, no two of these experiments use the same method to determine *G*. An uncertainty of only 14 p.p.m. is claimed by the University of Washington team[6] in Seattle (UWash-00), and this is still the record. This experiment[6] is elegant in both conception and execution. The University of Zurich group[7] (UZur-06) produced a remarkably similar result using a completely independent method. Problem solved? Not quite. The CODATA-06 error bars reflect the considerable scatter among the total set of *G* data that were considered[4] (only the two values with the smallest claimed uncertainties are shown in Fig. 2, and these results happen to agree perfectly). More recently, the group from the Huazhong University of Science and Technology in China (HUST-09) announced its final result[8]; and now, after extensive checking failed to uncover any errors in their work, Parks and Faller finally published[3] their *G* value (JILA-10).

Could something really fundamental be going on here? Probably not. It seems most unlikely that any discrepancies between different values could be due to a failure of classical physics to apply perfectly well to all of these experiments. Parks and Faller point out that their result agrees well with the CODATA-86 recommended value (Fig. 2). It is therefore interesting

that CODATA has previously explained why it considers the key datum in the 1986 analysis to have been superseded by later work[5].

Ironically, because the authors' experiment has no evident flaw, their measurement may lead CODATA to increase the uncertainty of its next recommended value of *G*. Stay tuned. ∎

Richard Davis *is at the Bureau International des Poids et Mesures, 92312 Sèvres Cedex, France.*
e-mail: rdavis@bipm.org

1. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics* Vol. 1, Ch. 7 (Addison-Wesley, 2005).
2. Gillies, G. T. *Rep. Prog. Phys.* **60,** 151–226 (1997).
3. Parks, H. V. & Faller, J. E. *Phys. Rev. Lett.* **105,** 110801 (2010).
4. Mohr, P. J., Taylor, B. N. & Newell, D. B. *Rev. Mod. Phys.* **80,** 633–730 (2008).
5. Mohr, P. J. & Taylor, B. N. *Rev. Mod. Phys.* **77,** 1–107 (2005).
6. www.npl.washington.edu/eotwash/experiments/bigG/bigG.html
7. Schlamminger, St. *et al. Phys. Rev. D* **74,** 082001 (2006).
8. Luo, J. *et al. Phys. Rev. Lett.* **102,** 240801 (2009).

EVOLUTIONARY GENOMICS

When abnormality is beneficial

One might think that aneuploidy — having an abnormal number of chromosomes — would be harmful, and would reduce an organism's fitness. Not necessarily: it all depends on the type of aneuploidy and the associated conditions. See Letter p.321

JUDITH BERMAN

Most cancer cells are aneuploid, and contain an unbalanced number of chromosomes. The question is whether aneuploidy can cause the unbridled growth of cancer cells, as Bovary proposed more than 100 years ago[1]. On page 321 of this issue, Pavelka *et al.*[2] demonstrate that different constellations of aneuploid chromosomes can confer a growth advantage on cells when they are exposed to stress conditions. This paper adds fuel to a long-standing controversy[3,4] over whether aneuploid chromosomes are good or bad for cell proliferation, and highlights the point that 'good' and 'bad' are relative terms that are highly dependent on the conditions under which they are measured.

Pavelka *et al.* exploited the facile genetics of the budding yeast *Saccharomyces cerevisiae* to produce a set of aneuploid strains in an unbiased manner. They first constructed strains that had three or five complete sets of whole chromosomes (triploids and pentaploids), instead of the usual two (diploids), and then induced them to undergo meiotic cell division. The odd numbers of starting chromosome sets ensured that a high frequency of spores would carry multiple aneuploidies. Other strengths of the study were the large number of genetically identical aneuploid strains generated (38), and the focus on strains that were stable and had undergone few cell divisions.

The authors analysed the progeny soon after birth — before single-nucleotide mutations could accumulate, as assessed by whole-genome deep sequencing. They found that most strains had decreased growth rates in nutrient-rich media, as well as under several stress conditions. Notably, however, most of the aneuploid strains grew faster than their parent strain on transfer to at least one stress condition, such as exposure to a chemotherapeutic agent or an antifungal drug. So it seems that some combinations of aneuploid chromosomes proliferate better under stress conditions, despite having had no prior exposure to that condition.

Intriguingly, in a number of cases, different constellations of aneuploid chromosomes conferred a similar growth advantage. This indicates that there is more than one way to get the job done. Whole-genome deep sequencing confirmed that no single-nucleotide mutations had accumulated in the five isolates Pavelka *et al.* analysed for messenger-RNA and protein composition; thus aneuploidy alone was sufficient to confer the growth advantages. The authors also show that an extra copy of a single gene, *ATR1*, conferred resistance to a tumorigenic compound. This bolsters previous observations in *S. cerevisiae*[5] and in another fungal species, *Candida albicans*[6], that aneuploidy can confer a selective advantage through the increased expression of one or more genes.

This paper, however, reaches different conclusions to those of Torres *et al.*[7,8], who reported that a single-nucleotide mutation in a deubiquitinating enzyme, which arose during the evolution of one aneuploid isolate, leads to improved proliferation of a few, but not all, strains. Pavelka *et al.*[2] find that aneuploidies alone — without any mutations — can confer improved growth under some stress

conditions. Both groups agree that, under conditions optimized by geneticists for growth in conventional laboratories, aneuploid cells usually divide less rapidly than cells with the normal chromosomal complement.

An important distinction in the methods used to generate the aneuploid strains might explain the differences in the findings[2,7,8]. Torres et al.[7,8] engineered yeast cells with a haploid (single) set of chromosomes to carry one extra chromosome and then selected for faster growth using conventional lab conditions for 9–14 days — a time frame during which mutations are expected to accumulate. By contrast, Pavelka et al. analysed strains that often carried multiple aneuploid chromosomes and, importantly, minimized the number of generations before analysis. This illustrates a crucial truism of experimental genetics: you get what you select for.

Pavelka and co-workers also directly address a controversy concerning the role of excess proteins in aneuploid cells. Previously, Torres et al.[7] proposed that there is a specific set of genes and proteins that are regulated in response to aneuploidy in general. In their more recent study[8], they showed that some 20% of proteins exhibit levels that do not track with gene copy number, and that a large proportion of these proteins are members of macromolecular complexes. By contrast, other groups[9,10] have found that the levels of most proteins generally reflect changes in chromosome copy number and that less than 5% of the proteins exhibit 'dosage compensation' — whereby the relative protein level is independent of gene-copy number. Pavelka et al. specifically test this hypothesis by quantitative mass spectrometry of about 2,000 proteins in each of five aneuploid strains and do not find compelling evidence for specific dosage compensation of protein-complex components.

Overall, these studies[2,7–10] are consistent with the idea that aneuploidy is not a single, unique state and that all aneuploid strains do not share a single, common phenotype or protein profile. Rather, different aneuploid strains use different mechanisms for optimal growth under different conditions. This conclusion may be less satisfying than a single, simple answer, especially given the crucial implications for cancer cells: it remains unclear whether cancer cells divide uncontrollably because they are aneuploid and/or because they have accumulated mutations that allow them to tolerate aneuploidy. But it should be remembered that work on cancer cells themselves[11] suggests that not all aneuploidies are equal: aneuploidy can either promote or inhibit tumorigenesis, depending on the context. Pavelka and colleagues' work[2] therefore supports the idea that, whereas mutations can facilitate the proliferation of aneuploid cells, aneuploidy itself can be sufficient to provide a growth advantage under a broad range of stress conditions. ■

Judith Berman *is in the Department of Genetics, Cell Biology, and Development, and the Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455, USA.*
e-mail: jberman@umn.edu

1. http://8e.devbio.com/article.php?ch=4&id=24
2. Pavelka, N. et al. Nature **468**, 321–325 (2010).
3. Duesberg, P., Li, R., Fabarius, A. & Hehlmann, R. Contrib. Microbiol. **13**, 16–44 (2006).
4. Hede, K. J. Natl Cancer Inst. **97**, 87–89 (2005).
5. Rancati, G. et al. Cell **135**, 879–893 (2008).
6. Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A. & Berman, J. Mol. Microbiol. **68**, 624–641 (2008).
7. Torres, E. M. et al. Science **317**, 916–924 (2007).
8. Torres, E. M. et al. Cell **143**, 71–83 (2010).
9. Geiger, T., Cox, J. & Mann, M. PLoS Genet. **6**, e1001090 (2010).
10. Springer, M., Weissman, J. S. & Kirschner, M. W. Mol. Syst. Biol. **6**, 368 (2010).
11. Weaver, B. A. A. & Cleveland, D. W. Cancer Res. **67**, 10103–10105 (2007).

HIGH-TEMPERATURE SUPERCONDUCTIVITY

Mind the pseudogap

The discovery of predicted collective electronic behaviour in copper–oxide superconductors in the non-superconducting state provides clues to unlocking the 24-year-old mystery of high-temperature superconductivity. SEE LETTER P.283

CHANDRA VARMA

The phenomenon of high-temperature superconductivity is a beautiful and well-posed scientific problem with many facets. On page 283 of this issue, Li et al.[1] report observing a special kind of intense collective electronic fluctuation in the most mysterious phase of matter exhibited by high-temperature superconducting copper-oxide materials (cuprates). Taken together with previous experimental[2–6] and theoretical[7] work, this observation significantly narrows the range of directions likely to be fruitful in the quest to understand high-temperature superconductivity. The authors performed their experiments in two samples of $HgBa_2CuO_{4+\delta}$, which has one of the simplest crystal structures of any of the cuprate families and is ideal for such studies.

Li and colleagues' experiments[1] pertain to the pseudogap region of the phase diagram of the cuprates (Fig. 1), a sort of precursor state to the superconducting phase that most condensed-matter physicists regard as the Rosetta Stone for discovering the physical principles that underlie the cuprates' behaviour. On entering the pseudogap region, at a temperature below T^* but above the temperature below which superconductivity emerges (T_c), all cuprates' thermodynamic and electronic-transport properties change by a large amount owing to the materials' loss of low-energy electronic excitations.

The pseudogap region is bounded on one side by a region of remarkably simple but unusual properties, which do not fit into the Fermi-liquid-type model that has been used to describe metals at low temperatures for about a hundred years. Some researchers got to grips with understanding this 'strange-metal' region early in the history of high-T_c superconductivity, by hypothesizing a quantum critical point in the dome-shaped superconductivity region of the phase diagram (Fig. 1). This point would occur at zero temperature and would involve a change in the symmetry of the materials' electronic structure. Because T_c is determined by the materials' collective electronic excitations in the non-superconducting state, it is unarguable that the coupling of electrons to such excitations in the strange-metal region and their modifications in the pseudogap region lead to high-T_c superconductivity.

If it exists, a quantum critical point in the
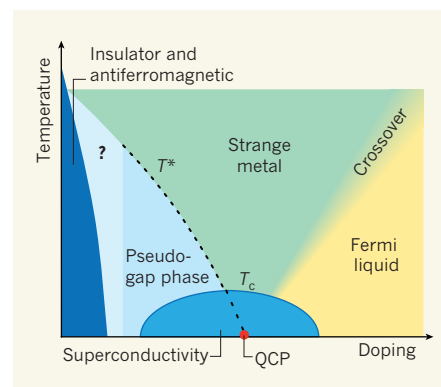
Figure 1 | Phase diagram of the cuprates. At very low levels of electron–hole doping, cuprates are insulating and antiferromagnetic (the materials' neighbouring spins point in opposite directions). At increased doping levels, they become conducting, and the exact temperature and doping level determine which phase of matter they will be in. At temperatures below T_c, they become superconducting, and at temperatures above T_c but below T^* they fall into the pseudogap phase. The boundary of the pseudogap region at low doping levels is unknown. The transition between the Fermi-liquid phase and the strange-metal phase occurs gradually (by crossover). QCP denotes the quantum critical point at which the temperature T^* goes to absolute zero. Li and colleagues' study[1] pertains to the pseudogap phase.

cuprates would also probably be the end point of a line of phase transitions that separates the pseudogap and strange-metal regions. However, the existence of such a line is a highly controversial issue, and justifiably so, because no singularity in the electronic properties — which is typically found in phase transitions — has been discovered at T^* despite the fact that more than 100,000 papers have been written on the subject. Most scientists have tended to believe that the pseudogap transition is a gradual crossover phenomenon rather than an abrupt phase transition. If it is a phase transition, the change in symmetry in the materials' electronic structure must be highly unusual and hard to discover, even though the 'order parameter' that characterizes the transition must be big enough to cause the large change observed in the materials' thermodynamic and transport properties.

Previous experimental work[2], based on a technique called polarized elastic neutron diffraction, demonstrated the emergence of an unusual long-range electronic order below about T^* in five samples of the yttrium–barium family of cuprates. This form of order was independently confirmed in another sample[3] of the yttrium–barium family, and also found in samples of three other cuprate families[4–6], including the mercury–barium class studied by Li and colleagues[1]. The type of symmetry associated with this order, deduced from these experiments[2–6], is consistent with a theoretical model[7] for the cuprates. This derives a long-range-ordered phase in which pairs of electron-current loops flow within each of the materials' unit cells and produce a pair of oppositely directed magnetic moments (Fig. 2). (In this model, at about T^*, a thermodynamic quantity, the specific heat, and some electronic-transport properties are expected to change smoothly.) Although these experiments[2–6] indicate that the pseudogap phenomenon is due to a genuine change in symmetry, it is crucial to have further experimental evidence of its signatures.

Given a long-range order, there must always be collective electronic fluctuations characteristic of that order. A cuprate in a state with loop-current order will condense into one of four possible ground-state (lowest-energy) configurations (Fig. 2). Collective fluctuations from such a condensed configuration can take the system to any of the three other arrangements (or to their linear combinations). Therefore, three modes of fluctuations, magnetic in nature and with finite energy at all momenta, are to be expected in the case of loop-current order.

Li *et al.* have observed one such mode of fluctuations in two samples of $HgBa_2CuO_{4+\delta}$, each with a different level of charge-carrier (electron–hole) doping, and confirmed the mode's magnetic nature through the technique of inelastic polarized neutron scattering. The authors have also found evidence for another
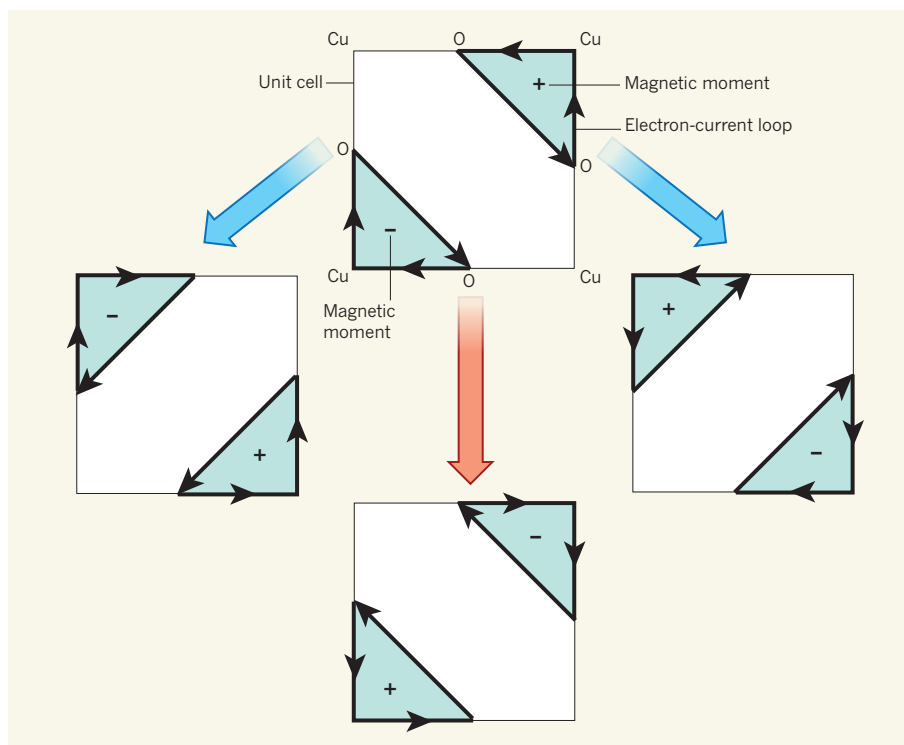
Figure 2 | Loop-current electronic order. A copper-oxide (CuO) material with loop-current electronic order has four possible ground-state configurations. In each configuration, pairs of electron-current loops flow within each of the material's unit cells and produce a pair of oppositely directed magnetic moments (the plus and minus symbols denote a magnetic moment with a direction that is perpendicular out of and into the plane of the page, respectively). If the material condenses into the top configuration, it can locally quantum-mechanically oscillate back and forth between it and the other three configurations. This gives rise to three possible collective modes of oscillation. Li and colleagues[1] have observed two of these modes.

of the three possible modes, but the identification of the third mode is beyond the reach of neutron-scattering techniques. They find that the intensity of the two observed modes begins to be noticeable at the temperature T^* of each of the samples (see Fig. 3a,b on page 285). At low temperatures, the combined intensity of the modes is consistent with the magnitude of the order observed in earlier experiments[4].

The significance of Li and colleagues' study[1] lies in confirming, through a quite different type of experiment, the discovery of a universal type of electronic order in the pseudogap regime of the cuprates. The modes observed are also the most likely candidates for the discrete modes that have been inferred indirectly in photoemission and infrared-absorption experiments[8]. The next step will be to understand whether such an electronic order and its fluctuations lead to all the unusual, universal properties of the cuprates. The expected quantum critical fluctuations of the strange-metal phase have already been derived on the basis of the observed order, and a theory of the coupling of such fluctuations to electrons has also been developed[9] to explain the properties of the strange-metal phase and the 'd-wave symmetry' of the materials' electronic structure in the superconducting phase.

However, the nature of the partial gap in the energy spectrum of the cuprates that

characterizes the pseudogap region itself remains to be understood. Only if it is understood can a valid theory, based on the observed long-range order and its fluctuations, be claimed for the cuprates' superconductivity. Achieving decisive consilience in this field will require various advances — understanding of the gap, further evidence of the order using other experimental techniques, confirmation of both of the collective modes observed by Li *et al.*[1] in other cuprates, and the discovery of the third mode. ∎

Chandra Varma *is in the Department of Physics and Astronomy, University of California Riverside, Riverside, California 92521, and is currently at Stanford University, Palo Alto, California 94305, USA. e-mail: chandra.varma@ucr.edu*

1. Li, Y. *et al. Nature* **468,** 283–285 (2010).
2. Fauqué, B. *et al. Phys. Rev. Lett.* **96,** 197001 (2006).
3. Mook, H. A., Sidis, Y., Fauqué, B., Balédent, V. & Bourges, P. *Phys. Rev. B* **78,** 020506 (2008).
4. Li, Y. *et al. Nature* **455,** 372–375 (2008).
5. Balédent, V. *Phys. Rev. Lett.* **105,** 027004 (2010).
6. Kaminski, A. *et al. Nature* **416,** 610–613 (2002).
7. Varma, C. M. *Phys. Rev. B* **55,** 14554–14580 (1997).
8. van Heumen, E. *et al. Phys. Rev. B* **79,** 184512 (2009).
9. Aji, V., Shekhter, A. & Varma, C. M. *Phys. Rev. B* **81,** 064515 (2010).

Rethinking schizophrenia

Thomas R. Insel[1]

How will we view schizophrenia in 2030? Schizophrenia today is a chronic, frequently disabling mental disorder that affects about one per cent of the world's population. After a century of studying schizophrenia, the cause of the disorder remains unknown. Treatments, especially pharmacological treatments, have been in wide use for nearly half a century, yet there is little evidence that these treatments have substantially improved outcomes for most people with schizophrenia. These current unsatisfactory outcomes may change as we approach schizophrenia as a neurodevelopmental disorder with psychosis as a late, potentially preventable stage of the illness. This 'rethinking' of schizophrenia as a neurodevelopmental disorder, which is profoundly different from the way we have seen this illness for the past century, yields new hope for prevention and cure over the next two decades.

The challenge of creating a vision of schizophrenia for 2030, which I attempt here, is a difficult one. There is certainly a risk in predicting scientific progress—the most important discoveries will probably be ones we cannot imagine today. But it is equally true that we can use past experience and the present state of knowledge to predict some aspects of the future. For schizophrenia, our knowledge base in 2010 is mostly based on clinical observation.

Schizophrenia is a syndrome: a collection of signs and symptoms of unknown aetiology, predominantly defined by observed signs of psychosis. In its most common form, schizophrenia presents with paranoid delusions and auditory hallucinations late in adolescence or in early adulthood. These manifestations of the disorder have changed little over the past century.

A century ago we had large public institutions for serious mental illness, tuberculosis and leprosy. Of these three, today only mental illness, especially schizophrenia, remains unchanged in prevalence and disability[1].

Sustained recovery occurs in less than 14% within the first five years following a psychotic episode[2]. Longer-term outcomes may be marginally better: a large international 25-year follow-up study reported an additional 16% with late-phase recovery[3]. Throughout Europe, less than 20% of people with schizophrenia are employed[4]. A large US study found nearly 20% homeless in a one-year follow up[5]. And a recent report from a patient advocacy group reported that in the US those with serious mental illness were three times more likely to be found in the criminal justice system than in hospitals. (http://www.treatmentadvocacycenter.org)

Although many have attributed this lack of progress to failed systems of care (http://www.mentalhealthcommission.gov/), we still do not have a basic understanding of the pathophysiology of the disorder and therefore lack the tools for curative treatment or prevention needed for most people with schizophrenia. If we are to transform outcomes by 2030, we can start by offering individuals and families challenged by serious mental illness a candid account of the current state of knowledge and a thoughtful consideration of future prospects.

One-hundred years of schizophrenia

The history of schizophrenia says more in many ways about the perspectives of the observer than the observed. In the late nineteenth century, Kraepelin defined "dementia praecox" or premature dementia as distinct from the insanity of tertiary syphilis or the cyclic, non-deteriorating psychosis of manic depressive illness[6]. Bleuler, who coined the term schizophrenia in the early twentieth century, was less convinced of its deteriorating course but emphasized the notion of a fundamental disorder of thought and feeling, which every psychiatrist for decades learned as the four 'a's—disturbances of associations, affect, ambivalence and autistic isolation[7]. These early formulations emerging before the split between neurology and psychiatry were consistent with the notion of a mental disorder rooted in brain pathology. Yet for much of the twentieth century, with the predominance of psychoanalytic theory, the study of the mind ignored the brain. Schizophrenia was a psychotic reaction, a fragmented ego due to a rejecting or ambivalent mother and treatments included re-mothering to build a stable ego[8].
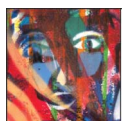
In the second half of the twentieth century, with the emergence of neuroleptic drugs, the pendulum swung in the other direction—a focus on brain chemistry deemphasized the mind. Schizophrenia was considered a 'dopamine disorder' based on the psychosis-inducing effects of dopamine-releasing drugs, such as amphetamine, and the anti-psychotic efficacy of a score of drugs that blocked the dopamine D2 receptor[9]. This neurochemical view of schizophrenia yielded medications that transformed the treatment of psychosis, allowing patients to be treated outside of hospitals and, in some cases, resulting in remission of the major symptoms of the illness. Early neuroleptic medications, examples of which are chlorpromazine and haloperidol, have been increasingly replaced by 'atypical' antipsychotics that have fewer extrapyramidal side effects (such as tremor and rigidity) but usually do not seem to be significantly more efficacious than the original dopamine D2 receptor antagonists[10]. Although both conventional and atypical antipsychotics reliably reduce delusions and hallucinations, they have not enhanced functional recovery (for example, employment) for people with schizophrenia. One explanation is that the disability of schizophrenia is largely due to cognitive deficits, such as problems with attention and working memory, which these drugs fail to improve.

A focus on cognitive symptoms has led to a more recent hypothesis of schizophrenia as a 'glutamate disorder' (reviewed in ref. 11) Healthy volunteers given low doses of NMDA receptor antagonists, such as ketamine, manifest select aspects of schizophrenia, including some of the attentional and memory problems. Conversely, agents that modulate the glycine modulatory site on the NMDA receptor have been reported to reduce some of the cognitive symptoms of schizophrenia. The theory is that schizophrenia, particularly the cognitive symptoms of the disorder, may result from low activity of the NMDA receptor on GABA inhibitory interneurons in the prefrontal cortex.

[1]National Institute of Mental Health, Bethesda, Maryland 20892, USA.

Although there can be little argument that medications have transformed the treatment of psychosis, research focusing on the drugs instead of the illness has thus far yielded too little progress on the pathophysiology of schizophrenia. It is not clear, for instance, that either dopamine D2 receptors or interneuron NMDA receptors are related to the cause of this disorder. Although post-mortem studies have consistently reported a loss of GABA and reductions in key enzymes for glutamate biosynthesis, potentially consistent with the glutamate hypothesis, these changes may represent the effects of chronic illness or treatment of the disorder rather than the cause of schizophrenia[11].

One approach that could separate cause from effect is genetics. Just as neuropharmacology dominated schizophrenia research in the late twentieth century, genetics has been a leading focus in the first decade of this century. Although in the 'genomic era' such a shift was inevitable, it was also presaged by a generation of twin and family studies demonstrating high heritability[12,13]. Reported concordance in monozygotic twins was roughly 50%, never the 100% figure one might expect for a Mendelian disorder, but considerably higher than dizygotic twins or siblings[14].

High heritability has not, however, translated into a satisfying search for genetic lesions. Although early genome-wide or candidate-gene studies searching for common variants associated with schizophrenia were mostly disappointing, either because early findings failed to replicate or large-scale studies failed to detect genome-wide significance, recent international consortia combining single nucleotide polymorphism (SNP) data from several independent studies have found replicable associations with genes of the major histocompatability complex (MHC) region on chromosome 6p21.3-22.1, *ZNF804A* on chromosomes 2q32.1, neuregulin 1 (*NRG1*) on chromosome 8, as well as transcription factor 4 (*TCF4*) on 18q21.2 (refs 15–17). Other studies have reported SNPs in candidate genes associated either with schizophrenia or a broad phenotype of psychosis, notably for genes within the neuregulin–ERBB4 signalling pathway[18], synaptic protein genes (for example, *NRX1* (also known as *PNO1*))[19], a potassium channel (*KCNH2*)[20] and many other brain-expressed proteins (for example, dysbindin)[21]. Currently, at least 43 candidate genes have been identified, but individual effect sizes are consistently modest (http://www.schizophreniaforum.org/res/sczgene/TopResults.asp), especially relative to the evidence for high heritability[22,23]. Epistatic or additive effects of these variants may explain more of the risk, but results thus far on individual variants from case–control studies have not been useful for understanding an individual's risk for schizophrenia.

In addition to the many reports of common single nucleotide variations, many rare structural genomic variants, such as copy number variants and translocations, have been described in schizophrenia (reviewed in ref. 24). These rare variants seem to have larger causative effects than previously reported SNPs, but most are not specific to schizophrenia and some occur only in a single family. The diversity and private nature of these mutations preclude a simple genetic explanation for schizophrenia, but these findings may yield important clues to pathophysiology. For instance, although the *DISC1* translocation that confers very high risk for psychiatric disorder has been detected in only a single Scottish family, this private mutation has revealed important mechanisms of disease and identified a site where common variation may also confer risk (reviewed in ref. 25). Even more encouraging, the consistent reports that so many of these structural variants affect genes implicated in brain development may predict the future of schizophrenia research.

Mapping the pathophysiology of schizophrenia

A starting point for mapping the pathophysiology of schizophrenia can begin with the increasing recognition that this is a neurodevelopmental disorder, or perhaps more accurately a collection of neurodevelopmental disorders that involve alterations in brain circuits. Although Feinberg[26], Weinberger[27] and Murray[28] proposed this approach more than two decades ago, the field is only now providing the evidence and recognizing the implications of shifting to a neurodevelopmental approach[29,30].

Psychosis nearly always emerges in late adolescence or early adulthood, with a peak between ages 18 and 25, when the prefrontal cortex is still developing. We still do not understand all of the changes in normal or abnormal cortical development during this period. Attempts to map functional connectivity defined by imaging the default network demonstrate little integration until after age nine[31]. Longitudinal neuroimaging studies demonstrate changes in grey matter density until the mid-twenties with the prefrontal cortex being the last to mature[32]. The cellular basis for the observed reduction in grey-matter density with magnetic resonance imaging (MRI) is not clear although classical anatomical post-mortem studies indicate that both synaptic elimination and increased myelination continue into early adulthood[33,34]. Whereas the literature from human post-mortem neuroanatomy of adolescence is scant, studies in non-human primate brain demonstrate that the refinement of circuits during early adulthood includes pruning of asymmetric (excitatory) synapses, proliferation of inhibitory circuits and the continued elaboration of pyramidal dendrites as targets of inhibitory input[35–37]. Together these observations indicate that this late stage of brain maturation involves a careful calibration of excitatory–inhibitory balance in the cortex with the prefrontal cortex the last region to mature (Fig. 1). As one potentially relevant modulator of this balance, dopamine innervation of the prefrontal cortex increases markedly during adolescence[38,39].

Although schizophrenic psychosis usually emerges between ages 18–25, several longitudinal population-based studies indicate that problems are evident much earlier. For instance, a recent report from a 45-year follow up of a Copenhagen birth cohort demonstrated that adults with schizophrenia have a history of delayed maturation including delayed developmental milestones in the first year[40]. Data from the Dunedin birth cohort, consistent with many previous studies[41], indicated that IQ is reduced early and persistently in children destined to develop schizophrenia[42]. These precursors of schizophrenia are subtle and non-specific, but the consistency of the finding supports the hypothesis that psychosis does not emerge from a completely healthy brain.

The emerging picture from genetic studies also indicates that early brain development is affected. As noted earlier, many of the structural variants associated with schizophrenia implicate neurodevelopmental genes involved with neuronal proliferation, migration, or synapse formation[43]. Even genes that are not exclusively developmental seem to influence schizophrenia by their early disruption[44]. In a particularly intriguing example, Niwa *et al.*[45] reported that a transient knockdown of *DISC1* in the frontal cortex in the pre- and perinatal mouse brain led to neurochemical and behavioural disruptions emerging in early adulthood. Moreover, some of the vulnerability alleles of candidate genes, such as *NRG1* and *DISC1*, seem to selectively influence splice variants expressed predominantly in developing cortex, implicating isoforms that show large developmental changes in expression in the prefrontal cortex[46–48]. As a final link to development, the genetics of schizophrenia overlaps with the genetics of autism and other neurodevelopmental disorders[19,49]. It is unclear why the same genetic variation associated with many different neurodevelopmental syndromes is manifested in some by age 3 years (autism) and in others after age 18 years (schizophrenia). Presumably there are genomic modifiers or possibly environmental influences that determine the specific syndrome. The study of discordant twins may yield important information for understanding the mismatches between genotype and phenotype.

Environmental factors identified so far have also implicated prenatal or perinatal events. Maternal malnutrition during famine[50,51], infections in the second trimester[52], perinatal injury[53] and cytokine exposure[54] have all been associated with subsequent increased risk for schizophrenia. Most of these effects are modest (less that twofold increase in risk) and none seem specific for schizophrenia, but in aggregate they demonstrate that early adverse experiences, including mid-gestational insults, are a risk factor for psychosis occurring two decades later. Gene-by-environment studies may demonstrate more robust effects[55], but an even more promising approach may be epigenetic maps indicating the 'scars' of early experience or the stochastic changes emerging across development[56]. As an example, a gene disrupted by a rare copy number variant in autism was found to be repressed by hypermethylation in a
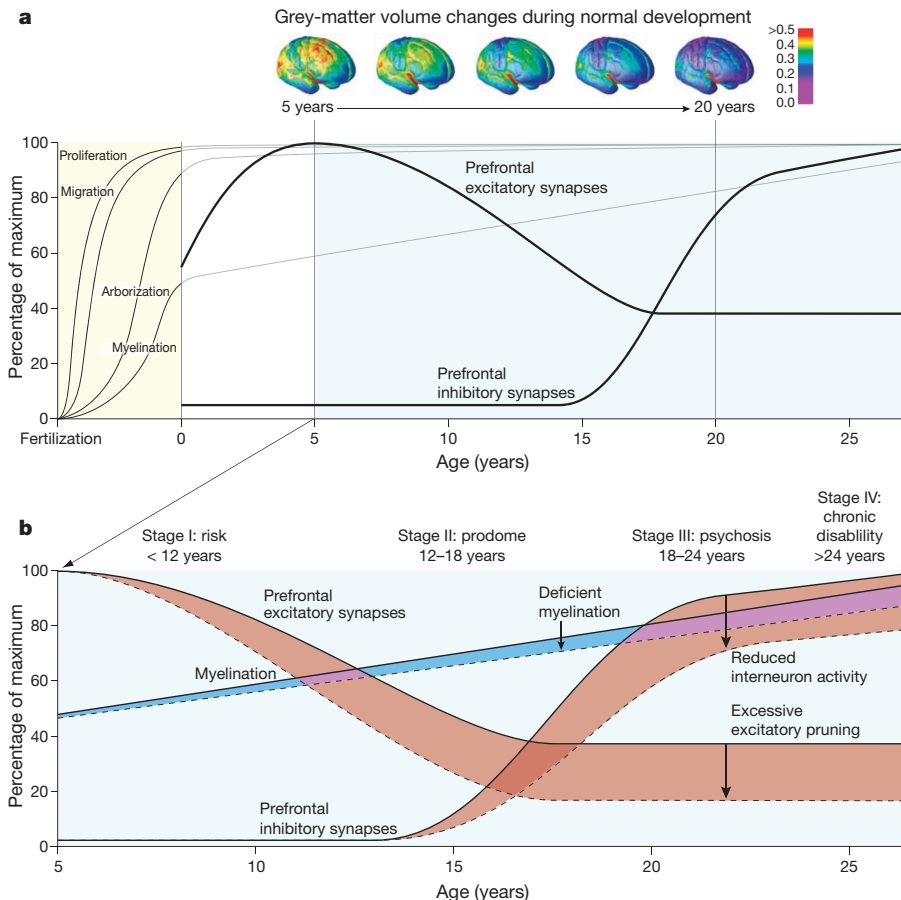
Figure 1 | Neurodevelopmental model of schizophrenia. a, Normal cortical development involves proliferation, migration, arborization (circuit formation) and myelination, with the first two processes occurring mostly during prenatal life and the latter two continuing through the first two post-natal decades. The combined effects of pruning of the neuronal arbor and myelin deposition are thought to account for the progressive reduction of grey-matter volume observed with longitudinal neuroimaging. Beneath this observed overall reduction, local changes are far more complex. Data from human and non-human primate brain indicate increases in inhibitory and decreases in excitatory synaptic strength occurring in prefrontal cortex throughout adolescence and early adulthood, during the period of prodrome and emergence of psychosis. **b**, The trajectory in children developing schizophrenia could include reduced elaboration of inhibitory pathways and excessive pruning of excitatory pathways leading to altered excitatory–inhibitory balance in the prefrontal cortex. Reduced myelination would alter connectivity. Although some data support each of these possible neurodevelopmental mechanisms for schizophrenia, none has been proven to cause the syndrome. Detection of prodromal neurodevelopmental changes could permit early intervention with potential prevention or preemption of psychosis.

large number of children with autism who had a perfectly normal genomic sequence[57].

The model that emerges from this neurodevelopmental perspective is that of an early insult, a latent period through much of neural development, and the emergence of psychosis in late adolescence or early adulthood. One possibility is a lesion early in development that does not manifest until a much later developmental stage when compensatory changes can no longer suffice. Thompson and Levitt[58] have called this developmental allostasis. A second, not mutually exclusive possibility is that the developmental lesion influences a pathway or a regulatory process, such as the fine tuning of excitatory and inhibitory synapses in the prefrontal cortex, which may have only subtle effects until a precise balance is required in late adolescence. Current data cannot distinguish between these two options, but either way a neurodevelopmental perspective implies the importance of timing and the opportunity for earlier intervention and prevention.

How will we map the trajectory of schizophrenia as a neurodevelopmental disease? Recent longitudinal studies of children with a rare, early-onset form of schizophrenia have used neuroimaging to identify differences in the trajectory of brain development. In these studies, children with schizophrenia seem to undergo excessive losses of grey matter and cortical thinning, essentially overshooting the normal pattern described earlier for adolescents[59,60]. These findings, although intriguing, are limited in that they do not reveal the changes before psychosis.

An opportunity for mapping earlier phases of the trajectory can be found in velocardiofacial syndrome, a syndrome associated with a micro-deletion of chromosome 22q11 (reviewed in ref. 61). Approximately 30% of children with a microdeletion of 22q11 will develop a form of schizophrenia that clinically and neurocognitively cannot be distinguished from the idiopathic disorder[62,63]. Most of these children are detected as toddlers because of their cardiac disease. Important insights into the trajectory from risk to disorder may be gained from ongoing longitudinal studies of these children comparing cognitive, affective and neural development in those who do and do not develop psychosis among this cohort with a similar genomic deletion.

Will animal studies reveal the neurodevelopmental trajectory of schizophrenia? Unlike the many disorders in medicine that can be modelled in mice or flies, an animal model of schizophrenia seems unlikely. Indeed, aspects of the prefrontal neuroanatomy and the executive function deficits of schizophrenia seem to be distinctively human. This is not to say that studies in animals, especially non-human primates, will be unimportant for schizophrenia. We lack fundamental information on the normal development of the forebrain, from the timing and geography of gene expression to the patterns of circuit formation under various environmental conditions. With current technology, these critical developmental maps will only be derived from studies in animals. Animal studies can also aid the study of abnormal development. Whereas animal models of schizophrenia are not likely, 'model animals' such as mice and flies

engineered with schizophrenia candidate genes will be highly informative for linking genetic variation to changes in cell and circuit function. For instance, mice with homologous deletions to the 22q11 lesion of velo-cardiofacial syndrome manifest differences in circuit formation and synaptic plasticity[64,65]. Such model animals will not only yield studies of disease mechanisms but opportunities for new treatment development.

Increasingly, however, it seems that humans may prove the best animal for modelling schizophrenia. Just as genes can create relevant models in non-human animals, genes can serve as a portal to mapping the pathophysiology of schizophrenia in cells from patients with the disorder. With induced pluripotent stem cells derived from fibroblasts of patients with schizophrenia, we should soon be able to study many different neural cell types, including their development, functional connections and response to perturbations. These cells do not need to recreate the disorder in a dish; they need only yield disordered molecular networks to reveal targets for developing new therapies. Through identifying new targets and high-throughput screening of existing small molecule libraries, we can expect the next generation of treatments for schizophrenia to be based on molecular pathophysiology rather than serendipity.

The stages of schizophrenia

Perhaps the most fundamental change from re-conceptualizing schizophrenia as a neurodevelopmental disorder is the notion of trajectory of illness. If the disorder begins in prenatal or perinatal life, then the psychosis of late adolescence must be seen not as the onset but as a late stage of the disorder. Indeed, we can begin to hypothesize four stages of schizophrenia, from risk to prodrome to psychosis to chronic disability[66] (Table 1). At present, the diagnosis is based on the symptoms and signs of psychosis. With the advent of biomarkers and new cognitive tools as well as the identification of subtle clinical features, we are beginning to detect earlier stages of risk and prodrome.[67]

The earliest stage is risk, before detectable deficits. In 2010 we do not have the risk architecture of this syndrome, but we can begin to see some of the outlines, based on genomics. Beyond the rare, highly penetrant mutations (for example, DISC1 and the 22q11 deletion), epistatic interactions between more common, less penetrant variations may yield higher predictions of risk than our current list. Of course, the 50% concordance rate of homozygous twins reminds us that genomics will not predict all forms of risk. Identifying environmental factors, detecting critical epigenomic modifications, or mapping neural circuit differences may render more of the blueprint for risk, much as the algorithms for coronary artery disease use family history, plasma lipids and dietary history. The extent to which the risk factors for schizophrenia will be modifiable in the sense that we can reduce the risk for coronary artery disease or lung cancer remains to be seen. And although this earliest stage may not involve distress or help-seeking, longitudinal studies have begun to identify subtle but reproducible evidence for behavioural and cognitive problems in early childhood[68–70]. To define this earliest stage we will need to define the full architecture of individual risk: genetic and epigenetic biomarkers, cognitive indicators and physiological predictors of vulnerability to the disorder.

Over the past two decades, the pioneering work of McGorry and his colleagues[71,72] has established the prodrome of schizophrenia as a valid second stage of the illness before psychosis. Whether defined as ultra-high risk or pre-psychosis, the prodrome is now identified based on changes in thoughts (for example, bizarre ideas falling short of psychotic ideation), social isolation and impaired functioning (for example, reduced school performance). Recognizing that these features might seem endemic to adolescence, the Structured Interview for Prodromal Syndromes (SIPS) was developed to distinguish high risk for psychosis from more common adolescent angst[73]. Recently a large multi-site project in the United States of 291 adolescents followed for 2.5 years reported that the prodrome represented a 405-fold increase in risk (relative to the general population) and that a combination of three factors (for example, genetic risk with recent functional decline, unusual thought content, and either suspicion/paranoia or reduced social functioning) resulted in a positive predictive power for conversion to psychosis of 74–81% (ref. 74). The addition of biomarkers, detected with functional or structural neuroimaging (reviewed in ref. 75), or the use of neuropsychological tests of reaction time or verbal memory[76,77] may enhance detection and increase the predictive power. Given the high rate of behavioural distress in adolescence and the likelihood that many with prodromal symptoms will either mature out of them or develop other disorders, the challenge is to increase sensitivity for detecting ultra-high risk while not sacrificing specificity[78]. Specificity is a challenge: many of those who seek help for prodromal symptoms will develop other forms of psychopathology, not schizophrenia. What will we need to define this stage of schizophrenia? Although standardized clinical assessments will help and longitudinal imaging may yield biomarkers, it is likely that cognitive changes, such as reductions in working memory, may be the best predictor of the psychotic phase of schizophrenia[79]. Over the next few years, cognitive neuroscience will have a critical role in providing the tools for increasing the sensitivity and specificity of the schizophrenic prodrome[80].

It is unclear to what extent intervening during the prodrome will either prevent or forestall psychosis. Results from single-site trials of atypical antipsychotics[81], antidepressants[82], lithium[83] and cognitive behaviour therapy[84] have had, at best, modest effects in reducing symptoms or preventing conversion to psychosis. A recent randomized double-blind placebo-controlled 12-week trial of long-chain omega-3 polyunsaturated fatty acids reported a 12-month conversion to psychosis in 2 of 41 (4.9%) individuals in the treated group versus 11 of 40 (27.5%) individuals in the placebo group[85]. Although promising, the overall rate of conversion (13 of 81) is lower than that observed in most prodromal cohorts. Current efforts to use cognitive remediation may identify a low-risk approach that could be used even if specificity were low[86]. An innovative, broad effort on youth mental health in Australia is addressing the issues of false positives, low specificity and potential stigma from early diagnosis by developing community-based, resilience-based interventions[66].

Stage III of schizophrenia is psychosis manifested by hallucinations, delusions, disorganization of thought and behaviour, and psychomotor abnormalities. It is now clear that negative symptoms (loss of will, anhedonia, poverty of thought) and cognitive deficits (reduced working

Table 1 | Stages of schizophrenia

	Stage I	Stage II	Stage III	Stage IV
Features	Genetic vulnerability Environmental exposure	Cognitive, behavioural and social deficits Help-seeking	Abnormal thought and behaviour Relapsing–remitting course	Loss of function Medical complications Incarceration
Diagnosis	Genetic sequence Family history	SIPS Cognitive assessment Imaging	Clinical interview Loss of insight	Clinical interview Loss of function
Disability	None/mild cognitive deficit	Change in school and social function	Acute loss of function Acute family distress	Chronic disability Unemployment Homelessness
Intervention	Unknown	Cognitive training? Polyunsaturated fatty acids? Family support?	Medication Psychosocial interventions	Medication Psychosocial interventions Rehabilitation services

Stage I, pre-symptomatic risk; stage II, pre-psychotic prodrome; stage III, acute psychosis; stage IV, chronic illness.

memory, poor cognitive control) are core features of the disorder that account for much of the long-term morbidity and poor functional outcomes[87]. Although the avolitional component of the disorder may define a special subgroup[88], there is a new consensus that the negative symptoms and cognitive aspects of pathology are major unmet therapeutic needs[89,90]. If risk is analogous to hyperlipidemia, prodrome comparable to angina, then psychosis can be thought of as myocardial infarction with frequent residual loss of function. In spite of consistently positive acute responses to antipsychotic medications and psychosocial treatments, relapse rates approach 80% (ref. 2). Cognitive deficits and negative symptoms, whether preceding or emerging with psychosis seem, at best, only modestly responsive to current antipsychotic treatments[91,92]. The most urgent research priority in the near term will be effective treatments for the cognitive deficits, including the lack of insight that often inhibits adherence to both medication and psychosocial treatments.

Stage IV of schizophrenia involves chronic disability. In 1988, in the height of the AIDS epidemic, the editor of *Nature* noted that "schizophrenia is arguably the worst disease affecting mankind, even AIDS not excepted."[93] Not all individuals progress to this late stage of the illness, but for those who do the disability is not only psychiatric but medical. The oft-cited psychiatric deficits lead to unemployment, homelessness and incarceration, as noted earlier. A Finnish birth cohort study recently reported a 7% rate of suicide in schizophrenia, accounting for 50% of all deaths by age 39 (ref. 94). The medical complications of chronic schizophrenia are less well known. In 2010, smoking and obesity are epidemic among people with schizophrenia, with estimates of nicotine dependence ranging from 58–90% (ref. 95) and metabolic syndrome (obesity, hyperlipidemia, hyperglycemia and hypertension) present in 40% (ref. 96). Life expectancy for those with serious mental illness has been estimated at 56 years, approximately 25 years of premature mortality resulting usually from cardiopulmonary disease or other chronic medical conditions[97]. Importantly, many of the medical complications of schizophrenia can be prevented through tobacco cessation, dietary management and programs to manage cardiovascular health.

Schizophrenia in 2030

What is the prognosis of schizophrenia for 2030? I will venture a few predictions based on hope more than knowledge and recognizing that progress in understanding and treating schizophrenia may come from distant fields of science that have not yet been engaged in this area (Fig. 2).

Prevention

Judging from the success of preventive approaches to cardiac death and disability, refocusing our approach to schizophrenia on early detection and early intervention could yield substantial improvements in outcomes over the next decade or two. This will, of course, require sensitive and specific diagnostic tools as well as safe and effective interventions. The diagnostic tools for schizophrenia, like the diagnostic tools for cardiovascular risk, will probably require a combination of approaches, including measures of genetic risk, imaging the efficiency of neural circuits, and, probably most specifically, early cognitive changes. Interventions that include an aggressive focus on cognition along with family support may prove surprisingly effective for preempting or forestalling psychosis. Although a 'statin-like' medication would be an unambiguous breakthrough, we should not assume that a medication will be more effective than harnessing the developing brain's intrinsic plasticity for reversing the neural trajectory that leads from risk to prodrome. If the preemptive interventions are as effective as what we have today for coronary artery disease and if these are widely deployed, by 2030 we should expect a profound reduction in first-episode psychosis.

Reducing the cognitive deficits

The disability of schizophrenia in 2010 results more from the underrecognized and treatment-refractory cognitive deficits than from the more obvious and frequently treatable positive symptoms[98]. Over the
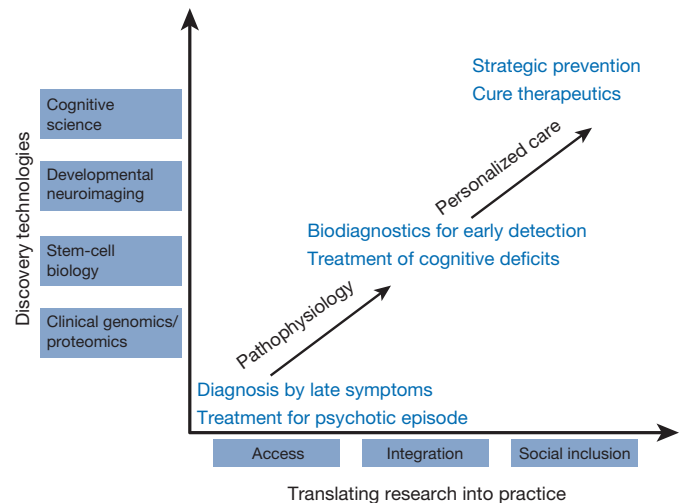
Figure 2 | A vision for schizophrenia over the next two decades. Currently diagnosis follows psychosis (stage III) and treatment focuses on reducing psychotic symptoms. The use of discovery technologies, which have already transformed the understanding and treatment of many other medical disorders, can transform our understanding of schizophrenia, yielding earlier diagnosis (stages I or II) with treatments focused on the cognitive deficits of this disorder. The ultimate goal, however, is cure and prevention based on an understanding of individual risk and the development of personalized care. In practice this means not only identifying risk and preemptive interventions but ensuring access to these interventions, integrating care and ensuring full social inclusion for people at any stage of the schizophrenia trajectory.

next decade, potentially leveraging current research on cognition in Alzheimer's disease, we can expect a series of pharmacological and nonpharmacological interventions that will reverse or mitigate the cognitive deficits of the disorder. Early initiation of these interventions will be transformative, but even in patients following psychosis, cognitive remediation may enhance employment, social inclusion and function in the community[99]. With interventions that reduce cognitive deficits, by 2030 we will be combining somatic, psychosocial and cognitive treatments with a goal of curing this disease for many patients.

Integration of care

One of the most egregious aspects of schizophrenia treatment in 2010 is the fragmentation of care, with medical care separated from psychiatric care and both isolated from psychosocial interventions, such as supportive employment and family education, which have a strong evidence base for effectiveness. Arguably, doing better with current treatments is our best short-term strategy for enhancing outcomes. A large multi-site effort in the United States, the Recovery After Initial Episode of Schizophrenia (RAISE) project, is developing a best-practices approach to bundled services that should provide some data about how much this can enhance outcomes. One can hope that in the near future, well before 2030, we will see all aspects of care being integrated in a continuous way, as is done increasingly for diabetes and other chronic disorders. Note, however, that the treatment of schizophrenia involves challenges not observed in most other chronic diseases. Denial of illness, paranoia, irrational thoughts, deficits in executive function and disruptive behaviour can all be part of the syndrome of untreated schizophrenia, complicating care for those with this disorder. Better treatments, not only better systems, will be necessary for better outcomes.

Stigma

Just as warehousing in institutions is mostly a memory today, imagine if the stigma associated with schizophrenia today were gone in 2030. In contrast to many other medical disorders, schizophrenia today too often defines a person rather than describing the illness. Our fear of psychosis or disruptive behaviour may keep us from seeing the heroic struggle that

people with this disorder face just to survive amidst the internal chaos and panic that is part of this chronic illness. Our expectations of these citizens are low: they should stay out of jail, on their medications and not distress their families, friends and fellow citizens. They deserve better. As a vision for 2030, people who suffer from any stage of schizophrenia will be considered to be educable, employable and capable of living in intimate relationships with others.

Will we still use the term schizophrenia in 2030? The accumulating genomic evidence indicates that there may be scores or hundreds of lesions contributing to this final common syndrome. The clinical evidence supports the possibility that what we have labelled schizophrenia for the past century may be many different disorders with different outcomes[88]. And the stigma associated with the diagnosis, and the past history of misunderstanding and mistreatment also indicate that a change in the term may be advisable. In 2002, the Japanese terms for schizophrenia 'Seishin-Bunretsu-Byo' ('mind-split disease') was replaced officially by 'Togo-Shitcho-Sho' ('integration disorder')[100]. Some evidence indicates that this name change led to reduced stigma, in that fewer people associated the new name with criminality[100].

Although semantic changes can be helpful, the transformations needed for those with this serious illness are likely to require not only a better label but better science (Fig. 2). In the next decade the challenge will be to integrate the impact of genetics, experience and development to identify a complete blueprint of the risk architecture of this syndrome. This should lead to a new taxonomy, identifying the many disorders within the syndrome we now call 'schizophrenia' and hopefully replacing this aggregate label with a series of more precise diagnoses based on pathophysiology. We need a personalized and preemptive approach, based on understanding and detecting individual risk and facilitated by safe and effective interventions for those in stages I and II of this disorder. In the meantime, we can create policies for social inclusion, family support and continuity of care to ensure that those in later stages of the syndrome have the best chance for recovery. Importantly, if recovery defined as a life in the community is our primary goal today, for 2030 our goals must include prevention, preemption and cure.

1. Hegarty, J. D., Baldessarini, R. J., Tohen, M., Waternaux, C. & Oepen, G. One hundred years of schizophrenia: a meta-analysis of the outcome literature. *Am. J. Psychiatry* **151**, 1409–1416 (1994).
2. Robinson, D. G., Woerner, M. G., McMeniman, M., Mendelowitz, A. & Bilder, R. M. Symptomatic and functional recovery from a first episode of schizophrenia or schizoaffective disorder. *Am. J. Psychiatry* **161**, 473–479 (2004).
3. Harrison, G. *et al.* Recovery from psychotic illness: a 15- and 25-year international follow-up study. *Br. J. Psychiatry* **178**, 506–517 (2001).
4. Marwaha, S. *et al.* Rates and correlates of employment in people with schizophrenia in the UK, France and Germany. *Br. J. Psychiatry* **191**, 30–37 (2007).
5. Folsom, D. P. *et al.* Prevalence and risk factors for homelessness and utilization of mental health services among 10,340 patients with serious mental illness in a large public mental health system. *Am. J. Psychiatry* **162**, 370–376 (2005).
6. Kraepelin, E. *Dementia Praecox and Paraphrenia* (Chicago Medical Book Co., 1919).
7. Bleuler, E. *Dementia Praecox or the Group of Schizophrenias* (International Universities Press, 1950).
8. Arieti, S. *Interpretation of Schizophrenia* (Basic Books, 1974).
9. Carlsson, A. The current status of the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology* **1**, 179–186 (1988).
10. Lieberman, J. A. *et al.* Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N. Engl. J. Med.* **353**, 1209–1223 (2005).
    Clinical results of the largest publically funded treatment trial for chronic schizophrenia demonstrates no significant advantage for second-generation anti-psychotic medications, except for clozaril.
11. Coyle, J. T. Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cell. Mol. Neurobiol.* **26**, 363–382 (2006).
12. Kety, S. S. The significance of genetic factors in the etiology of schizophrenia: results from the national study of adoptees in Denmark. *J. Psychiatr. Res.* **21**, 423–429 (1987).
13. McGuffin, P. & Gottesman, I. I. Risk factors for schizophrenia. *N. Engl. J. Med.* **341**, 370–372 (1999).
14. Cardno, A. G. *et al.* Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch. Gen. Psychiatry* **56**, 162–168 (1999).
15. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
    Largest meta-analysis of genomic variants in schizophrenia demonstrates genome-wide significance for three common variants as well as several rare copy number variants.
16. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
    These results demonstrate similar genetic risk architecture for schizophrenia and bipolar disorder, indicating that the genetics of psychosis may not be specific to any of the currently identified psychiatric syndromes.
17. Ettinger, U. *et al.* Prefrontal and striatal volumes in monozygotic twins concordant and discordant for schizophrenia. *Schizophr. Bull.* doi: 10.1093/schbul/sbq060 (10 June 2010).
18. Hahn, C. G. *et al.* Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nature Med.* **12**, 824–828 (2006).
19. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–996 (2009).
20. Huffaker, S. J. *et al.* A primate-specific, brain isoform of *KCNH2* affects cortical physiology, cognition, neuronal repolarization and risk of schizophrenia. *Nature Med.* **15**, 509–518 (2009).
21. Straub, R. E. *et al.* Genetic variation in the 6p22.3 gene *DTNBP1*, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am. J. Hum. Genet.* **71**, 337–348 (2002).
22. McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210–217 (2010).
23. Need, A. C. *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet.* **5**, e1000373 (2009).
24. Bassett, A. S., Scherer, S. W. & Brzustowicz, L. M. Copy number variations in schizophrenia: critical review and new perspectives on concepts of genetics and disease. *Am. J. Psychiatry* **167**, 899–914 (2010).
25. Brandon, N. J. *et al.* Understanding the role of DISC1 in psychiatric disease and during normal development. *J. Neurosci.* **29**, 12768–12775 (2009).
26. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17**, 319–334 (1982).
27. Weinberger, D. R. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44**, 660–669 (1987).
28. Murray, R. M., Jones, P. & O'Callaghan, E. Fetal brain development and later schizophrenia. *Ciba Found. Symp.* **156**, 155–170 (1991).
29. Lewis, D. A. & Levitt, P. Schizophrenia as a disorder of neurodevelopment. *Annu. Rev. Neurosci.* **25**, 409–432 (2002).
30. Jaaro-Peled, H. *et al.* Neurodevelopmental mechanisms of schizophrenia: understanding disturbed postnatal brain maturation through neuregulin-1–ErbB4 and DISC1. *Trends Neurosci.* **32**, 485–495 (2009).
31. Fair, D. A. *et al.* The maturing architecture of the brain's default network. *Proc. Natl Acad. Sci. USA* **105**, 4028–4032 (2008).
32. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge during adolescence? *Nature Rev. Neurosci.* **9**, 947–957 (2008).
33. Huttenlocher, P. R. Synapse elimination and plasticity in developing human cerebral cortex. *Am. J. Ment. Defic.* **88**, 488–496 (1984).
34. Yakovlev, P. & Lecours, A. R. in *Regional Development of the Brain in Early Life* (ed. Minkowski, A.) (Blackwell Scientific, 1967).
35. Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N. & Goldman-Rakic, P. S. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232–235 (1986).
36. Hashimoto, T. *et al.* Protracted developmental trajectories of GABA$_A$ receptor α1 and α2 subunit expression in primate prefrontal cortex. *Biol. Psychiatry* **65**, 1015–1023 (2009).
37. Lewis, D. A. & Gonzalez-Burgos, G. Neuroplasticity of neocortical circuits in schizophrenia. *Neuropsychopharmacology* **33**, 141–165 (2008).
38. Rosenberg, D. R. & Lewis, D. A. Postnatal maturation of the dopaminergic innervation of monkey prefrontal and motor cortices: a tyrosine hydroxylase immunohistochemical analysis. *J. Comp. Neurol.* **358**, 383–400 (1995).
39. Lambe, E. K., Krimer, L. S. & Goldman-Rakic, P. S. Differential postnatal development of catecholamine and serotonin inputs to identified neurons in prefrontal cortex of rhesus monkey. *J. Neurosci.* **20**, 8780–8787 (2000).
40. Sørensen, H. J. *et al.* Early developmental milestones and risk of schizophrenia: a 45-year follow-up of the Copenhagen Perinatal Cohort. *Schizophr. Res.* **118**, 41–47 (2010).
41. Woodberry, K. A., Giuliano, A. J. & Seidman, L. J. Premorbid IQ in schizophrenia: a meta-analytic review. *Am. J. Psychiatry* **165**, 579–587 (2008).
42. Reichenberg, A. *et al.* Static and dynamic cognitive deficits in childhood preceding adult schizophrenia: a 30-year study. *Am. J. Psychiatry* **167**, 160–169 (2010).
43. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
44. Li, W. *et al.* Specific developmental disruption of disrupted-in-schizophrenia-1 function results in schizophrenia-related phenotypes in mice. *Proc. Natl Acad. Sci. USA* **104**, 18280–18285 (2007).
45. Niwa, M. *et al.* Knockdown of DISC1 by *in utero* gene transfer disturbs postnatal dopaminergic maturation in the frontal cortex and leads to adult behavioral deficits. *Neuron* **65**, 480–489 (2010).
    Mouse study uses transient knockdown of DISC1 in forebrain to show the unique developmental effect of this protein, with biochemical, physiological and behavioural changes emerging after puberty.
46. Tan, W. *et al.* Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. *J. Biol. Chem.* **282**, 24343–24351 (2007).
47. Colantuoni, C. *et al.* Age-related changes in the expression of schizophrenia susceptibility genes in the human prefrontal cortex. *Brain Struct. Funct.* **213**, 255–271 (2008).

48. Nakata, K. et al. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. Proc. Natl Acad. Sci. USA 106, 15873–15878 (2009).
Study of DISC1 shows isoforms expressed predominantly during human brain development are influenced by alleles associated with schizophrenia.

49. Guilmatre, A. et al. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch. Gen. Psychiatry 66, 947–956 (2009).

50. Susser, E. S. & Lin, S. P. Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944–1945. Arch. Gen. Psychiatry 49, 983–988 (1992).

51. St Clair, D. et al. Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959–1961. J. Am. Med. Assoc. 294, 557–562 (2005).

52. Brown, A. S. & Derkits, E. J. Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. Am. J. Psychiatry 167, 261–280 (2010).

53. Cannon, M., Jones, P. B. & Murray, R. M. Obstetric complications and schizophrenia: historical and meta-analytic review. Am. J. Psychiatry 159, 1080–1092 (2002).

54. Ellman, L. M. et al. Structural brain alterations in schizophrenia following fetal exposure to the inflammatory cytokine interleukin-8. Schizophr. Res. 121, 46–54 (2010).

55. Nicodemus, K. K. et al. Serious obstetric complications interact with hypoxia-regulated/vascular-expression genes to influence schizophrenia risk. Mol. Psychiatry 13, 873–877 (2008).

56. Feinberg, A. P. & Irizarry, R. A. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc. Natl Acad. Sci. USA 107 (suppl. 1), 1757–1764 (2010).

57. Gregory, S. G. et al. Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med. 7, 62 (2009).
Novel observation of hypermethylation of candidate gene in autism indicates potential mechanism by which rare structural variants can reveal sites for epigenetic modification leading to changes in gene expression.

58. Thompson, B. L. & Levitt, P. Now you see it, now you don't—closing in on allostasis and developmental basis of psychiatric disorders. Neuron 65, 437–439 (2010).

59. Gogtay, N. et al. Comparison of progressive cortical gray matter loss in childhood-onset schizophrenia with that in childhood-onset atypical psychoses. Arch. Gen. Psychiatry 61, 17–22 (2004).
First demonstration of the specificity of cortical changes in childhood-onset schizophrenia, providing evidence for excessive loss of gray matter in prefrontal cortex.

60. Rapoport, J. L. et al. Progressive cortical change during adolescence in childhood-onset schizophrenia. A longitudinal magnetic resonance imaging study. Arch. Gen. Psychiatry 56, 649–654 (1999).

61. Karayiorgou, M., Simon, T. J. & Gogos, J. A. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. Nature Rev. Neurosci. 11, 402–416 (2010).
Important, comprehensive review of basic and clinical research on 22q11 microdeletions, providing a pathway for linking genetic lesions to the pathophysiology of schizophrenia.

62. Bassett, A. S. et al. The schizophrenia phenotype in 22q11 deletion syndrome. Am. J. Psychiatry 160, 1580–1586 (2003).

63. Chow, E. W., Watson, M., Young, D. A. & Bassett, A. S. Neurocognitive profile in 22q11 deletion syndrome and schizophrenia. Schizophr. Res. 87, 270–278 (2006).

64. Mukai, J. et al. Palmitoylation-dependent neurodevelopmental deficits in a mouse model of 22q11 microdeletion. Nature Neurosci. 11, 1302–1310 (2008).

65. Stark, K. L. et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nature Genet. 40, 751–760 (2008).

66. McGorry, P. D., Yung, A. R., Bechdolf, A. & Amminger, P. Back to the future: predicting and reshaping the course of psychotic disorder. Arch. Gen. Psychiatry 65, 25–27 (2008).

67. Nestler, E. J. & Hyman, S. E. Animal models of neuropsychiatric disorders. 13, 1161–1169 (2010).

68. Cannon, T. D. et al. Childhood cognitive functioning in schizophrenia patients and their unaffected siblings: a prospective cohort study. Schizophr. Bull. 26, 379–393 (2000).

69. Cornblatt, B., Obuchowski, M., Roberts, S., Pollack, S. & Erlenmeyer-Kimling, L. Cognitive and behavioral precursors of schizophrenia. Dev. Psychopathol. 11, 487–508 (1999).

70. Erlenmeyer-Kimling, L. et al. Attention, memory, and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High-Risk Project. Am. J. Psychiatry 157, 1416–1422 (2000).
Classic study identifying childhood predictors of adult psychosis in longitudinal design.

71. Henry, L. P. et al. The EPPIC follow-up study of first-episode psychosis: longer-term clinical and functional outcome 7 years after index admission. J. Clin. Psychiatry 71, 716–728 (2010).

72. Yung, A. R. et al. Validation of ''prodromal'' criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. Schizophr. Res. 105, 10–17 (2008).

73. Woods, S. W. et al. Validity of the prodromal risk syndrome for first psychosis: findings from the North American Prodrome Longitudinal Study. Schizophr. Bull. 35, 894–908 (2009).

74. Cannon, T. D. et al. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. Arch. Gen. Psychiatry 65, 28–37 (2008).
Multi-site study to validate the prodrome of schizophrenia, identifying predictors of conversion to psychosis.

75. Smieskova, R. et al. Neuroimaging predictors of transition to psychosis—a systematic review and meta-analysis. Neurosci. Biobehav. Rev. 34, 1207–1222 (2010).

76. Cohen, J. D., Barch, D. M., Carter, C. & Servan-Schreiber, D. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. J. Abnorm. Psychol. 108, 120–133 (1999).

77. Seidman, L. J. et al. Neuropsychology of the prodrome to psychosis in the NAPLS consortium: relationship to family history and conversion to psychosis. Arch. Gen. Psychiatry 67, 578–588 (2010).

78. Riecher-Rössler, A. et al. Efficacy of using cognitive status in predicting psychosis: a 7-year follow-up. Biol. Psychiatry 66, 1023–1030 (2009).

79. Kremen, W. S. et al. Cognitive decline in schizophrenia from childhood to midlife: a 33-year longitudinal birth cohort study. Schizophr. Res. 118, 1–5 (2010).

80. Carter, C. S., Barch, D. M., Gur, R., Pinkham, A. & Ochsner, K. CNTRICS final task selection: social cognitive and affective neuroscience-based measures. Schizophr. Bull. 35, 153–162 (2009).

81. Woods, S. W. et al. Randomized trial of olanzapine versus placebo in the symptomatic acute treatment of the schizophrenic prodrome. Biol. Psychiatry 54, 453–464 (2003).

82. Cornblatt, B. A. et al. Can antidepressants be used to treat the schizophrenia prodrome? Results of a prospective, naturalistic treatment study of adolescents. J. Clin. Psychiatry 68, 546–557 (2007).

83. Berger, G. et al. Neuroprotection in emerging psychotic disorders. Early Interv. Psychiatry 1, 114–127 (2007).

84. Morrison, A. P. et al. Cognitive therapy for the prevention of psychosis in people at ultra-high risk: randomised controlled trial. Br. J. Psychiatry 185, 291–297 (2004).

85. Amminger, G. P. et al. Long-chain omega-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. Arch. Gen. Psychiatry 67, 146–154 (2010).

86. Keefe, R. S. et al. Report from the Working Group Conference on multisite trial design for cognitive remediation in schizophrenia. Schizophr. Bull. doi:10.1093/schbul/sbq010 (1 March 2010).

87. Hyman, S. E. & Fenton, W. S. Medicine. What are the right targets for psychopharmacology? Science 299, 350–351 (2003).

88. Kirkpatrick, B., Buchanan, R. W., Ross, D. E. & Carpenter, W. T. Jr. A separate disease within the syndrome of schizophrenia. Arch. Gen. Psychiatry 58, 165–171 (2001).

89. Buchanan, R. W. et al. A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. Schizophr. Bull. 31, 5–19 (2005).

90. Kirkpatrick, B., Fenton, W. S., Carpenter, W. T. Jr & Marder, S. R. The NIMH-MATRICS consensus statement on negative symptoms. Schizophr. Bull. 32, 214–219 (2006).

91. Keefe, R. S. et al. Effects of olanzapine, quetiapine, and risperidone on neurocognitive function in early psychosis: a randomized, double-blind 52-week comparison. Am. J. Psychiatry 164, 1061–1071 (2007).

92. Swartz, M. S. et al. Effects of antipsychotic medications on psychosocial functioning in patients with chronic schizophrenia: findings from the NIMH CATIE study. Am. J. Psychiatry 164, 428–436 (2007).

93. Editor. Where next with psychiatric illness? Nature 336, 95–96 (1988).

94. Alaräisänen, A. et al. Suicide rate in schizophrenia in the Northern Finland 1966 Birth Cohort. Soc. Psychiatry Psychiatr. Epidemiol. 44, 1107–1110 (2009).

95. Kelly, D. L. et al. Cigarette smoking and mortality risk in people with schizophrenia. Schizophr. Bull. doi:10.1093/schbul/sbp152 (17 December 2009).

96. McEvoy, J. P. et al. Prevalence of the metabolic syndrome in patients with schizophrenia: baseline results from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial and comparison with national estimates from NHANES III. Schizophr. Res. 80, 19–32 (2005).

97. Colton, C. W. & Manderscheid, R. W. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. Prev. Chronic Dis. 3, A42 (2006).

98. Harvey, P. D., Reichenberg, A., Bowie, C. R., Patterson, T. L. & Heaton, R. K. The course of neuropsychological performance and functional capacity in older patients with schizophrenia: influences of previous history of long-term institutional stay. Biol. Psychiatry 67, 933–939 (2010).

99. Bell, M. D., Zito, W., Greig, T. & Wexler, B. E. Neurocognitive enhancement therapy with vocational services: work outcomes at two-year follow-up. Schizophr. Res. 105, 18–29 (2008).

100. Takahashi, H. et al. Impact of changing the Japanese term for ''schizophrenia'' for reasons of stereotypical beliefs of schizophrenia in Japanese youth. Schizophr. Res. 112, 149–152 (2009).

PERSPECTIVE

From maps to mechanisms through neuroimaging of schizophrenia
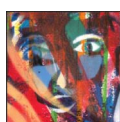
Andreas Meyer-Lindenberg[1]

Functional and structural brain imaging has identified neural and neurotransmitter systems involved in schizophrenia and their link to cognitive and behavioural disturbances such as psychosis. Mapping such abnormalities in patients, however, cannot fully capture the strong neurodevelopmental component of schizophrenia that pre-dates manifest illness. A recent strategy to address this issue has been to focus on mechanisms of disease risk. Imaging genetics techniques have made it possible to define neural systems that mediate heritable risk linked to candidate and genome-wide-supported common variants, and mechanisms for environmental risk and gene–environment interactions are emerging. Characterizing the neural risk architecture of schizophrenia provides a translational research strategy for future treatments.

One-hundred-and-two years after the term schizophrenia was first used, this disorder remains one of the most serious, disabling and baffling brain diseases. Modern imaging techniques (Fig. 1) have been very useful in mapping out networks in the brain that are affected in patients with schizophrenia. This has contributed to establishing schizophrenia research firmly in the broader neuroscience community, and has the potential to result in reduced stigma for patients and their families. However, a main hope of advancing the neuroscience of schizophrenia is that this will lead to new and better treatments, which are needed urgently[1]. For this goal, mapping is not enough: convergent evidence shows that the disease process of schizophrenia long precedes manifest illness[2], and that abnormalities found in patients may reflect a complex and advanced condition that could be too late in the trajectory of the disease for guiding causal treatment or prevention, similar to being able to diagnose coronary artery disease only at the point of myocardial infarction. Can systems-level neuroscience help towards advancing the translational enterprise? Here, I propose ways in which this might be possible. After briefly recapitulating known functional, structural and network abnormalities in schizophrenia, I will discuss data indicating that neuroimaging is indeed useful in characterizing both genetic and environmental risk factors for schizophrenia that are likely to be causally related to the illness. I finish with some suggestions on how characterized mechanisms of illness risk can be effectively interfaced with traditional translational and drug-development processes.

Brain regions involved in schizophrenia

Structure

Extensive work studying the neuroanatomy of schizophrenia using imaging has shown clear abnormalities in at-risk subjects and both first-episode and chronic patients (Fig. 2). Overall grey matter, white matter and whole brain volume are decreased, whereas ventricular volume is increased. In the beginning of the disease, volumes are decreased in the hippocampus, thalamus, the left uncus/amygdala region, the bilateral insula and the anterior cingulate[3]. In chronic schizophrenia, more extensive volume reductions are observed in the cortex, particularly in medial and left dorsolateral prefrontal cortex, but also in the left superior temporal gyrus[3]. The magnitude of these alterations is mostly small to moderate, and there is considerable overlap between patient and control distributions. Hippocampal volume reductions are found in relatives of schizophrenia patients[4,5], indicating a heritable component, whereas the evidence for disease-related heritability for other cortical and subcortical features is more mixed[5]. Volume increases in first-episode schizophrenia are restricted to parts of the putamen and spread in chronic schizophrenia throughout the dorsal striatum[3]. These increases are not heritable[5] and are probably a consequence of antipsychotic drug action. In addition to volume changes, abnormalities in cortical thickness, gyrification and subcortical shapes have been reported.

Microcircuits

Corresponding to these macroscopic alterations in the brain of schizophrenia patients are changes in local microcircuits[6] (Fig. 3). In prefrontal cortex, pyramidal neurons—the main source of excitatory cortical–cortical neurotransmission—are reduced in size and packed more densely[7], indicating a reduction in axon terminals and dendritic spines that occupy
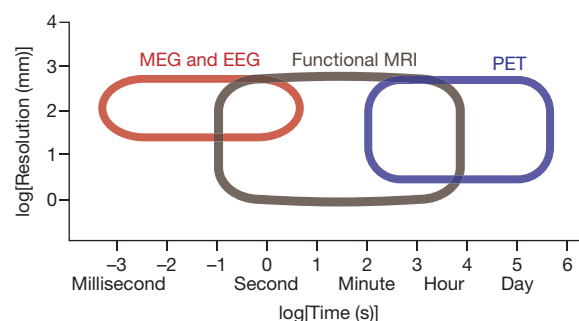
Figure 1 | Functional neuroimaging methods and their temporal and spatial resolution. Magnetoencephalography (MEG) and electroencephalography (EEG) image the electromagnetic effects of neuronal (assembly) action; their temporal resolution can be on the order of milliseconds whereas their spatial resolution tends to be less than that of fMRI, which images blood flow or oxygenation effects of neuronal activation, and PET, which uses radioisotopes to label molecules in the brain. fMRI and PET, in turn, are limited in their temporal resolution to several 100 ms (for fMRI) and minutes (for PET).

---

[1]Central Institute of Mental Health, University of Heidelberg/Medical Faculty Mannheim, J5, 68159 Mannheim, Germany.

cannabinoid receptor CB1. In the hippocampus, cell bodies of pyramidal neurons are smaller, dendritic spines are reduced, and there is inconsistent evidence of aberrantly located or clustered neurons in adjacent structures, especially the endorhinal cortex. In the thalamus, some studies indicate reductions in neuron number, especially in the mediodorsal nucleus and pulvinar[9]. Taken together, these findings are in reasonable agreement with structural imaging results. They suggest abnormalities in local processing, especially in glutamatergic drive to GABAergic parvalbumin-containing interneurons and intracortical connectivity, but are also indicative of changes in long-range connectivity, including thalamic afferents. An important source of modulation in prefrontal cortex function is dopaminergic. The delicate balance between information maintenance and flexible adjustment of information that characterizes executive function depends critically on an optimal level of dopamine signalling[10] which reaches prefrontal cortex from midbrain and ventral striatum; these prefrontal inputs appear to be reduced in schizophrenia.

Function

Cortical and subcortical information processing is functionally abnormal in both first-episode and chronic schizophrenia (Fig. 1). Most functional studies use an 'activation paradigm' in which a cognitive task is used to engage brain systems of interest, and the results are therefore usefully summarized under these cognitive domains, without suggesting a causal relationship between cognitive (sub)function, brain system and schizophrenia symptoms.

Executive function. Much attention has been focused on executive function (including working memory and selective attention), subserving flexible adaptation of behavioural patterns to external demands. Here, patients show quantitative abnormalities in dorsolateral prefrontal cortex (which have been linked to negative symptoms[11]), rostral anterior cingulate and inferior parietal lobule. In dorsolateral prefrontal cortex, patients show relatively inefficient prefrontal activation under low cognitive load, indicative of decreased signal-to-noise ratio, and a decrease in activation when executive demands exceed capacity[12]. There is evidence for compensatory activation in the ventrolateral prefrontal

Figure 2 | Brain regions functionally and/or structurally affected in schizophrenia. Modified, with permission, from ref. 6. PFC, prefrontal cortex.

the space between neurons that may be a consequence of exuberant synaptic pruning during adolescence. Several interneuron populations in the prefrontal cortex are reduced, such as those containing parvalbumin, which show consistent signs of reduced GABAergic neural transmission[8], but also those expressing neuropeptides such as somatostatin or the

Figure 3 | Schematic summary of putative alterations in dorsolateral prefrontal cortex circuitry in schizophrenia. Modified, with permission, from ref. 6. Grey, cortical pyramidal neuron; green, parvalbumin-containing interneuron; purple, basket neuron; red, thalamic neuron; yellow, dopaminergic neuron in brainstem.

cortex in patients, indicating a system that comes 'online' as the dorso-lateral-prefrontal and anterior cingulate system starts to fail[13]. Before manifest illness, vulnerability to psychosis has been associated with abnormal prefrontal activation at an intermediate level in high-risk populations[14].

Episodic memory. Episodic memory depends on interactions between the hippocampal formation and regions of the prefrontal cortex. In schizophrenia patients, dorsolateral prefrontal cortex activation is abnormally decreased[15]. Less consistently[14,15], decreased activation has been found in the hippocampal formation, possibly because it is part of the 'resting state network' and therefore it is difficult to achieve a reliable baseline.

Reward and salience. In reward and salience processing, prediction errors are signalled by midbrain dopamine neurons projecting to the ventral striatum and dorsolateral prefrontal cortex. For perceptual salience tasks, increases of midbrain and ventral striatal signals have been demonstrated in schizophrenia[16] and in high-risk subjects. Opposed to this, ventral striatal responses to reward seem to be reduced in schizophrenia in several[17], but not all, studies[18] and correlate with negative symptoms[17].

Emotional regulation. In the domain of emotional regulation, activation of amygdala to emotional images seems to be consistently reduced in schizophrenia patients[19], but not relatives[20], whereas neutral face expressions may lead to greater limbic activation in schizophrenia patients and is correlated with flat affect[21]. In a circuit between medial prefrontal cortex and amygdala, which is critical for the regulation of emotion processing, schizophrenia patients, but not healthy relatives, show reduced functional connectivity[20].

Social cognition. An area that has recently come under increased focus is social cognition. Neuroimaging has identified abnormalities in the medial prefrontal cortex, the temporoparietal junction and the amygdala in schizophrenia[22]. Prefrontal abnormalities are consistent with the interpretation that patients 'hyper-mentalize' (that is, show prefrontal activation for stimuli that have no objective social or intentional content[23]), a possible mechanism leading to delusions.

Hallucinations. One of the principal symptoms of schizophrenia is hallucinations, especially the perception of voices in the absence of external stimuli. Imaging results have demonstrated activity of auditory and speech processing cortices during hallucinatory experiences[24]. These findings seem to correlate with the extent of functional and structural connectivity abnormalities to speech areas in the temporal lobe[25], lending support to the idea that dysconnectivity of this region is important.

Connectivity

Since Wernicke's proposal at the end of the nineteenth century a disturbance of integrated activity has been viewed as fundamental for schizophrenia. Neuroimaging has been useful in defining abnormal circuits, especially with dorsolateral prefrontal cortex, and has shown that, rather than a uniform disruption or disconnectivity, schizophrenia is characterized by 'dysconnectivity': functional interactions are altered in a regionally and functionally differentiated manner.

During working memory, dorsolateral prefrontal cortex connectivity is altered in patients with schizophrenia and subjects at risk[26,27]. Interhemispheric prefrontal connectivity is reduced in patients and relatives, whereas a dysfunctional increase in the connectivity with the hippocampal formation (Fig. 4a) has been observed in chronic[28] and first-episode psychosis, and in at-risk subjects[29]. Data from resting-state networks indicate that dysfunctional increases of connectivity may be found within the extended limbic system in patients and subjects at risk[27]. A similar lateral-neocortical versus temporal and extended limbic distinction is suggested by multivariate analyses[30,31]. Recently, methods from topology have begun to be applied to brain networks[32]. One conclusion emerging from this work is that the human brain has properties in common with other complex systems (such as the Internet) that support an efficient and robust transfer of information while keeping wiring between regions low[32]. These 'small world' properties may be altered in schizophrenia[33] and predict impaired cognitive performance. Abnormalities of adult brain network organization related to schizophrenia would be expected to follow aberrant early brain development, but we know little about the normal development of brain functional networks or how this might be perturbed pre-clinically in individuals at high risk. Further signatures of abnormal local processing and connectivity defined using
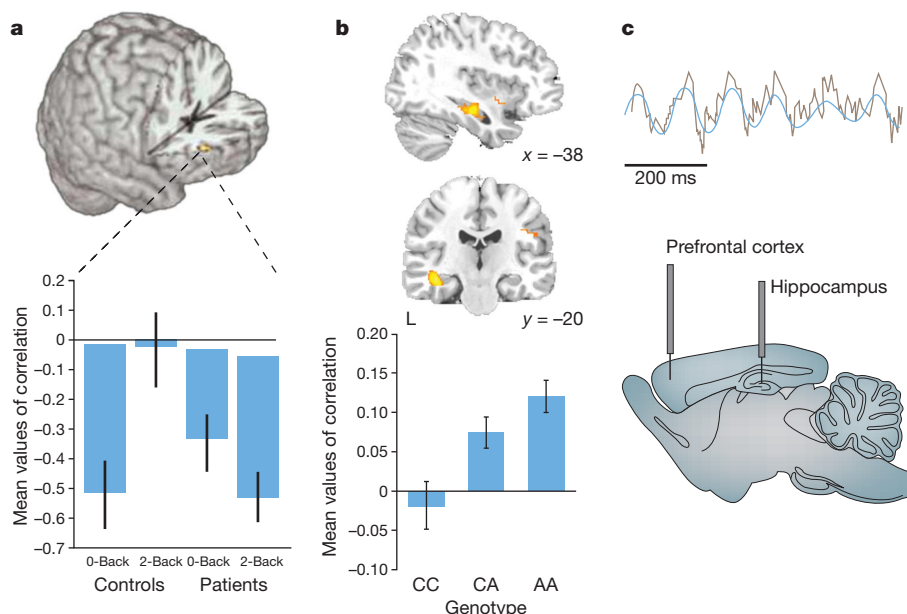
Figure 4 | A systems-level phenotype in patients relates to genetic risk and animal models. a–c, Abnormal prefrontal–hippocampal connectivity (measured as correlation of activity in PET across task conditions) during working memory (2-back) compared to a control task (0-back) in patients with schizophrenia and matched controls (modified, with permission, from ref. 28) (**a**), in carriers of the genome-wide significant genetic risk variant (genotype AA) in *ZNF804A* (modified, with permission, from ref. 76), which again shows persistent coupling between prefrontal cortex and hippocampal formation, this time measured with fMRI during the n-back working memory task (**b**), and in electrophysiological measurements in a mouse model of high genetic risk (22q11DS) (modified, with permission, from ref. 84) (**c**). Error bars, standard error.

functional magnetic resonance imaging (fMRI) and positron-emission tomography (PET) can be found with electrophysiological imaging (see Box 1).

Of these abnormal functional dorsolateral prefrontal cortex interactions, two systems deserve further discussion. The hippocampal formation provides input to the dorsolateral prefrontal cortex, and neonatal hippocampal formation lesions in animals induce prefrontal cortex abnormalities post-pubertally[34], indicating a causal role of the interaction between these two regions in schizophrenia. This fronto-hippocampal dysconnection hypothesis is also attractive as the hippocampal formation is selectively vulnerable to some early neurodevelopmental disturbances[2], such as obstetrical insults. Second, multiple parallel interactions between prefrontal cortex, thalamus and striatum form feedback loops critical for basic information processing; these feedback loops are disturbed in schizophrenia patients[35]. This prefrontal-neostriatal system is modulated by midbrain dopaminergic neurons which project to cortex and striatum and are, in turn, regulated by prefrontal cortex efferents[36]. This system is relevant for an understanding of acute psychosis.

A mechanistic account of acute psychosis

Abnormal dopaminergic neutrotransmission is known to be important for psychosis because the effectiveness of antipsychotic drugs is directly related to dopamine D2 receptor blockade[37]. Therefore, psychosis has been linked to a 'hyperdopaminergic' state, a hypothesis later modified to posit increases in striatum, while cortical dopamine was supposed to be reduced[38]. Unambiguous evidence for this from neuroimaging is only provided for striatum, where patients with schizophrenia have elevated presynaptic dopamine synthesis and dopamine release and moderate increases of dopamine D2/3 receptor levels[39]. Notably, elevated dopamine synthesis and release is also seen in subjects at risk for schizophrenia and in the prodromal state[40], indicating that they are part of the risk architecture of the illness.

Although assays of dopamine D1 and D2/3 receptors in prefrontal cortex have remained inconclusive, there is evidence that functional activation abnormalities in prefrontal cortex are tightly linked to striatal dopaminergic disinhibition in schizophrenia patients[41] and high-risk subjects[42], indicating an abnormality of prefrontal regulation of the midbrain-striatal dopamine system. Because dopaminergic signals in midbrain and striatum are essential for the signalling of salience, abnormal subcortical dopamine release could lead to aberrant assignment of salience to non-salient events, providing a plausible account for the emergence of psychosis[43]. Although functional imaging evidence for this model has been provided[16], it is indirect because measurements of midbrain dopamine release in relationship to psychosis in patients with schizophrenia have not yet been performed. Further work will also be necessary to clarify how other aspects of acute psychosis (such as hallucinations) are linked to dopamine dysregulation, as a salience account is unlikely to explain the entire spectrum of positive symptoms.

Why mapping is not enough

Despite these successes in delineating abnormalities in schizophrenia across the lifespan, such findings in manifestly ill patients can be related to numerous confounds other than illness status. For example, patients with schizophrenia often smoke, take a variety of medications, are in poorer somatic health, have different lifestyles and educational socio-economic trajectories, may have been frequently hospitalized, and so on. Because it is almost never possible to control for these confounding measures, the true causal contribution of a systems-level finding obtained in this way is never certain. Even more importantly, as the evidence supporting the neurodevelopmental hypothesis of schizophrenia indicates, data gathered during the stage at which schizophrenia is currently defined capture a stage of the illness that may characterize the brain at too late a stage for intervention, and thus do not offer much in terms of finding new, and especially causal, treatments. Given the clear evidence from heritability studies that a large proportion of schizophrenia risk is related to genes, and a smaller, but still sizable, proportion is

---

BOX 1

Oscillations and schizophrenia

Oscillations are important organizers of brain activity, plasticity and connectivity[100]. They can be measured using electroencephalography (EEG) or magnetoencephalography (MEG). Oscillations in the gamma range are important for synchronized activity within local cortical networks. Essential for the generation of local gamma activity are parvalbumin-containing GABAergic interneurons under glutamatergic stimulation[101]. Cognition requires that the results of local computations are globally integrated. Neural oscillations in the low (especially theta) ranges are critical for long-range connectivity because they engage larger areas and effectively modulate fast local oscillations, such as gamma oscillations[102]. In hippocampus, highly synchronized theta frequency oscillations are observed which have been proposed to serve as a temporal organizer for cortex[103]. This has recently been demonstrated in mouse[104], where hippocampal theta oscillations drive cortically generated gamma oscillations through phase locking. Importantly, NMDA (N-methyl-D-aspartate) antagonism can influence both local and long-range synchronization[105] because NMDA receptors in superficial layers of cortex, the main recipients of long cortical connections, control local processing. Dopamine modulates these oscillations[106]. In the prefrontal cortex in schizophrenia patients, reductions in the gamma and theta band have been observed at rest and during stimulus processing[107]. Aspects of these features seem to be present in first-degree relatives of patients with schizophrenia, indicating a role in the risk architecture[108].

Temporal coordination of oscillatory activity is critical for experience-dependent plasticity and therefore in the maturation of cortical networks. For spike-timing-dependent plasticity to occur, a window of the order of tenths of a millisecond for the co-occurrence of pre- and postsynaptic spiking has been proposed[109], which can be achieved through co-stimulation of cortical neurons over the theta-cycle of the hippocampus. This opens the possibility that aberrant oscillations during critical periods can have an enduring effect on the shaping of cortical circuits beyond their immediate impact on local processing. Compromising both long-range coupling (through white matter tract maldevelopment or lesions) and local processing (for example, in interneurons) could have enduring effects on synaptic plasticity. Dopamine could have a modulating role in this process because intact mesocortical dopaminergic input is necessary for long-term potentiation to occur at hippocampal–prefrontal cortex synapses[110], reflecting dopamine-D2-receptor-mediated dopaminergic control over NMDA-receptor-dependent synaptic plasticity in prefrontal cortex, indicative of a 'gating function' of dopamine D2 receptors. A further link to the neurodevelopmental hypothesis is provided by the observation that long-range synchronization of the theta and gamma band undergoes profound changes during adolescence[111], when cortical–cortical connectivity continues to mature through myelinization of long-range tracks. This indicates that the reduction of transmission delays between brain regions during adolescence, especially between hippocampus and prefrontal cortex, enables the kind of precise temporal coordination that is important for activity-dependent shaping of prefrontal circuits. Importantly, this emergence of long-range connectivity has been linked to maturation of cortical grey matter[112], indicating a causal sequence. Speculatively, in the context of the interaction of hippocampus and prefrontal cortex, a sequence of events seems credible in which hippocampal dysfunction leads to abnormal shaping of neurocortical circuits as soon as hippocampal–prefrontal connections become sufficiently stable during late adolescence. This deficit could even become progressive if experience-dependent plasticity continues throughout adult life.

related to environmental risk factors (as well as to the interaction between genes and environment), this indicates a strategy in which systems-level neuroscience is used to interrogate the neural effects of identified risk factors on the brain in an attempt to define a neural risk architecture of the illness. Because both common genetic and environmental risk factors affect healthy subjects as well as patients, such studies can often avoid the confounders associated with manifest illness and offer the hope to identify mechanisms that lie before the emergence of frank psychosis. Imaging genetics, which combines structural and functional imaging with genetic characterization of healthy participant samples, has shown itself to be a sensitive[44,45] and specific method to define such mechanisms.

Recent genome-wide evidence indicates that many thousands of genetic variants explain a sizeable proportion of genetic risk for schizophrenia[46]. This high complexity leads to challenges for imaging genetics, which uses the methods of genetic association with brain phenotypes. Just as one variant can have pleiotropic effects, so can several different genes influence the same neural pathways to risk. Results may be influenced by the genetic background. Although imaging genetics studies have provided evidence for both pleiotropy[45] and epistasis, the problem of interacting genetic variants remains difficult. Although a variety of studies have investigated two or three[47] variant interactions, few have been replicated so far and the underlying complexity is probably higher. This may be addressed by emerging multivariate methods that can deal with a large number of single nucleotide polymorphisms (SNPs) together with complex brain phenotypes[48], but these still need to be validated. Attention also needs to be paid to the heritability and reliability of the imaging paradigms.

As the 'endophenotype' concept[49] predicts, the penetrance of genetic effects on the level of brain imaging is high: two meta-analyses of imaging genetics[44,45] found effect sizes of 0.7–1.0, very considerably higher than what was found for association of the same variants with psychiatric diagnoses[50]. In addition, imaging genetics has the critical advantage of mapping genetic effects across the brain, in many cases allowing researchers to tie in the large body of preclinical knowledge that specifies how a given neural system—affected by genetic variant—functions in healthy subjects and what molecular, cell biological and systems-level factors influence its development and neural processing.

Risk mechanisms in schizophrenia
Candidate genes
Because single common risk variants for schizophrenia only cause moderate increases in relative illness risk[46,51,52], it is not surprising that association evidence is often inconsistent, especially as the current definition of schizophrenia, which is based on patient introspection and clinical observation, is unlikely to correspond to one well-defined biological entity. Also, with regards to genetics, not enough functional variants are yet known in genes of interest, and functional genomics approaches are necessary to identify them, especially in very large genes such neuregulin 1 (NRG1). Nevertheless, several such variants have repeatedly found support in association studies and are backed up by meta-analysis, justifying their investigation through systems-level neuroscience techniques. Such systems-level findings, in turn, can serve as one approach to in vivo functional genomics that can aid the discovery of functional variants. In the following brief overview, I cannot cover the range of candidate genes explored using imaging genetics and schizophrenia. Therefore, I will provide three examples: catechol-O-methyltransferase (COMT), NRG1 and disrupted in schizophrenia 1 (DISC1). They are typical candidate genes in the sense that association with the disease phenotype has been variable (meaning that they are not, strictly speaking, unambiguous schizophrenia-associated genes) despite pronounced systems-level and cognitive effects.

COMT has been the most-studied gene in the imaging genetics literature[53]. It encodes an enzyme that degrades catecholamines, including dopamine. COMT is particularly concentrated in the extrasynaptic spaces of the prefrontal cortex and hippocampus. Because prefrontal dopamine transporters are scarce, COMT is thought to have a key role in clearing dopamine in the prefrontal cortex[54]. An evolutionarily recent functional SNP in COMT results in the amino acid substitution of valine (Val) with methionine (Met) at codon 158 (rs4680), leading to a significant decrease in enzymatic activity in the brain and lymphocytes[55] of the Met allele, which therefore causes a higher level of prefrontal extracellular dopamine. The functional literature on the common rs4680 Val/Met polymorphism in COMT shows a highly consistent and large effect of rs4680 on prefrontal activation[45]. Effects of rs4680 on brain structure are less consistent, possibly because they may differ in directionality between prefrontal cortex and hippocampus[56] and show significant interactions with another putatively functional promoter region SNP. In multimodal neuroimaging, rs4680 modulated the functional interactions between midbrain dopamine synthesis and prefrontal function[57], mirroring post-mortem findings and indicating an entry point into the neural circuit for acute psychosis described above through this genetic risk variant.

NRG1 was first implicated in schizophrenia in an Icelandic sample[58]. NRG1 and its receptor ERB4 have important functions during brain development through signalling axon guidance, progenitor cell proliferation and neural migration in cortex, and seem to have a special role in shaping the development of parvalbumin-containing GABAergic interneurons[59]. Postnatally, NRG1 is implicated in activity-dependent plasticity at glutamatergic synapses, myelination and oligodendrocyte differentiation[60]. Neuroimaging has uncovered possible functional and structural correlates of dysmaturation associated with genetic variants in this system. In high-risk individuals, carriers of a NRG1 risk SNP had an increased risk for psychosis, compromised activation in medial prefrontal and temporooccipital regions during a sentence completion task, as well as impaired prefrontal and middle temporal lobe activation during semantic fluency[61]. Hippocampal volumes were smaller in carriers of a risk haplotype of NRG1 (ref. 62), and a risk SNP in the same region was associated in patients with larger ventricular volumes[63]. That same SNP also associated with reduced structural connectivity in healthy controls studied with diffusion tensor imaging[64].

DISC1 was implicated by the discovery of a translocation disrupting the gene in a large Scottish pedigree with a high density of mental disorders[65]. DISC1 is a multifunctional anchoring molecule that regulates different subcellular compartments, including at the synapse[66]. It is involved in neural progenitor proliferation, differentiation and radial migration and dendritic arborization[66]. In adult brain, DISC1 is highly expressed in hippocampus, where it has a key role in regulating adult neurogenesis.

Neuroimaging has identified the effects of genetic risk variants in DISC1 and prefrontal and hippocampal structure, function and interactions. A functional Ser704Cys polymorphism (Ser substituted for Cys at position 704) impacts on hippocampal structure and function[67], and prefrontal efficiency during verbal fluency[68]. Hippocampal formation–dorsolateral prefrontal cortex functional connectivity was increased[69] in risk allele carriers, an intermediate connectivity phenotype similar to that seen in overt disease (Fig. 4a). A common haplotype was associated with reduced grey matter in hippocampus and more prominently in prefrontal cortex[70].

It is interesting to consider possible molecular points of convergence between these candidate risk gene systems[71]. It has previously been noted that multiple candidate genes have an impact on the plasticity of glutamatergic synapses[72]. The recent evidence reviewed above extends this conclusion into the domain of early brain development. Both ERB4 and DISC1 are located in the postsynaptic density of glutamatergic synapses[73], where they co-localize with other susceptibility factors for schizophrenia and are exposed to varying levels of extraneuronal dopamine regulated by COMT. Furthermore, activity-dependent synaptic pruning is likely to be mediated by all of these factors. Neuroimaging data are beginning to define functional interactions between these risk variants and the impact they have on prefrontal cortex activity and brain structure[74], validating these ideas from cellular

neuroscience on the systems level. It will be important to examine the neural circuits so defined in new animal models that carry several of these genetic risk variants, permitting an examination of their convergence on pre- and postnatal maturation and synaptic pruning, especially in adolescence.

Genome–wide supported variants

Despite their clear impact on imaging phenotypes[45], the usefulness of candidate genes for understanding schizophrenia is a subject of debate because their association with the categorical disease phenotype itself is inconsistent. Genome-wide association studies (GWAS) offer an alternative, hypothesis-free way to identify genetic variants associated with schizophrenia. This is especially welcome when treatment implications are considered, for which one needs to study factors clearly related to risk. Although GWAS will probably not provide all of the answers about the genetics of schizophrenia, any common variant that does survive the extreme amount of statistical thresholding that this method requires certainly merits study using intermediate imaging phenotypes[46,51,52]. Of those variants, the one with the strongest support is zinc finger protein 804A (ZNF804A)[75], encoding a protein of unknown, but possibly regulatory, function. Like many candidate gene variants, ZNF804A is pleiotropic on the level of psychiatric diagnoses, also being associated with bipolar disorder[75]. In functional neuroimaging with a so-called 'n-back' working memory probe[76], healthy carriers of risk genotypes exhibited no changes in regional activity. However, they did exhibit pronounced gene-dosage-dependent alterations in functional connectivity, which was decreased from dorsolateral prefrontal cortex across hemispheres and increased with hippocampus (Fig. 4b), as described above for schizophrenia patients. Subsequent work has further implicated this variant in cognitive performance for executive cognition and episodic memory specifically[77], highlighting domains that are especially dependent on prefrontal–hippocampal interactions. Impaired white matter volume and integrity markers in carriers of this risk variant have been observed[78], as well as the inability to downregulate key parts of the mentalizing system in conjunction with impaired connectivity of this system to dorsolateral prefrontal cortex[79], indicating possible structural substrates and downstream functional activation effects of impaired prefrontal connectivity that mirror findings in patients[23]. Interestingly, abnormally increased coupling of amygdala was also observed, a phenotype unlikely to be related to heritable risk for schizophrenia[20] and therefore possibly related to risk for bipolar disorder, where similar findings in patients have been described. Another instructive variant is near calcium channel, voltage-dependent, L type, alpha 1C subunit (CACNA1C), first discovered in a GWAS for bipolar disorder[80] but subsequently implicated in schizophrenia. Healthy carriers of this variant showed impaired hippocampal activation and connectivity during episodic memory[81], mirroring findings in overt schizophrenia, as well as abnormalities in subgenual cingulate and amygdala[82], highlighting a key regulatory system for emotion and affect implicated in affective disorders. Because bipolar disorder and schizophrenia share a large proportion of genetic risk[46], it is noteworthy that both CACNA1C and ZNF804A have an impact on circuits that support a pleiotropic effect on both disorders. Further work should study the remaining group of currently established SNPs[46,51,52] with genome-wide significance for schizophrenia, including a variant upstream of neurogranin (NRGN) and a SNP in transcription factor 4 (TCF4), both probably involved in brain development, as well as a cluster in the major histocompatibility complex region on chromosome 6p22.1, which could indicate a gene by environment interaction system by implicating the immunological system in the pathogenesis of schizophrenia.

Microdeletions

As reviewed elsewhere, one important finding from GWAS is the increased occurrence of structural variations (microdeletions or microduplications) in schizophrenia, but possibly not with bipolar disorder. Of these, only 22q11, causing velocardiofacial or 22q11DS syndrome, was known previously. The new microdeletions are not solely associated with schizophrenia but also with other brain phenotypes such as mental disability, epilepsy and autism[83]. Despite these pleiotropic phenotypical effects and their relative rarity, which makes them account for only a minority of disease risk, identifying and characterizing structural variations holds considerable potential because each of these are associated with significant risk, which exceeds that from common genetic variants[84]. Although none of the newly identified variants has been characterized on the systems level, previous work on the 22q11 deletion[84] shows that a multimodal imaging approach is feasible and holds the promise to identify neural systems-level abnormalities associated with high genetic risk. Furthermore, as it seems likely that microdeletion risk cannot be explained by deletion or duplication of single genes, but rather interactions of genes jointly affected in their expression[85], imaging genetics can ask whether variants in such genes converge on neural systems implicated in schizophrenia. An example of this is the 22q11 microdeletion, which includes, besides COMT and several other candidate genes for schizophrenia, the proline oxidase gene (PRODH). A recent neuroimaging study showed that functional polymorphisms in PRODH associated with schizophrenia risk had an impact on prefrontal connectivity[86].

Environmental risk mechanisms in human brain

As reviewed elsewhere, convergent evidence supports an effect of environmental risk factors such as urban birth, prenatal stress, childhood trauma, migration and high expressed emotion. Although, as a group, these risk factors explain less risk than genes, individually they have an associated risk that far exceeds that of common genetic variants. The mechanisms underlying these environmental factors are largely unknown and are unlikely to be unitary. It has been proposed that one such mechanism may be social stress that plays out through activation of the hypothalamic–pituitary–adrenal axis and dopaminergic sensitization[87]. One salient feature of social stress that has been hypothesized to underlie schizophrenia is social defeat[88], defined as a subordinate position or outsider status, especially if repeatedly experienced. Although direct epidemiological evidence for this hypothesis is missing, experimental studies provide a link to key neural systems features implicated in risk for schizophrenia. In animals, social defeat stress increases the firing rate of dopaminergic neurons in midbrain area and brain-derived neurotrophic factor (BDNF)-dependent activity in the ventral striatum. This indicates links to neural plasticity for which BDNF, a gene inconsistently associated with schizophrenia, is essential, and to the pathophysiology for psychosis outlined above. Although this specific finding has not been established in humans, acute psychosocial stress[89] evokes striatal dopamine release, measured by PET. A further link to environmental risk factors is provided by the observation that in individuals with low maternal care, dopamine release was stronger[89], as in subjects with schizophrenia-associated personality characteristics[90]. Although there is indicative data linking social stress in general to subcortical dopamine systems, it remains a potentially fruitful, but currently almost unexplored, application of systems-level neuroscience to define mechanisms of specific environmental risk factors.

An early attempt in this direction has addressed the neural processing of stable and unstable social hierarchies[91]. Social status strongly predicts well being, morbidity and survival. Patients with schizophrenia are strongly over-represented in the lower social strata. In a fMRI experiment, unstable hierarchies, which are associated with health risk, showed specific recruitment of, among others, amygdala and medial prefrontal cortex[91], identifying a key regulatory system for the processing of negative affect that has been associated with genetic risk factors for affective disorders and schizophrenia, for example in CACNA1C (ref. 81), and linking status processing to key theory-of-mind regions impaired in schizophrenia[22] and carriers of risk genes in ZNF804A[79]. It remains to be seen if amygdala and regulatory systems in the medial prefrontal cortex are also associated with other social risk factors such as migration and urbanicity.

Identifying such neuroenvironmental risk systems will also advance the field of gene–environment interactions on the systems level. For example, it has been observed that carriers of the rs4680 Val risk variant of *COMT* have higher risk for psychosis when exposed to cannabis[92]. Neuroimaging has begun to delineate a mechanism by showing that dopamine release in prefrontal cortex—measured indirectly via PET—by the psychotogenic component of cannabis, Δ9-tetrahydrocannabinol, is modulated by this *COMT* risk variant[93]. Further demonstration of gene–environment interactions will become feasible as genetic and environmental 'main effect' brain mechanisms become better defined and identified.

Systems–level strategies for translation in schizophrenia

The ultimate goal of defining risk mechanisms for schizophrenia preceding frank psychosis is prevention. For this, the tools described above and the dysfunctional circuits that they have defined will have to be applied in large cohorts longitudinally to define transition likelihoods and intervention points. Research of this kind is now underway (for example, in the European IMAGEN study). However, we cannot wait for these data to come in before acting, as translational research in schizophrenia is in urgent need of a conceptual redesign. There is no evidence that the excess mortality of schizophrenia has decreased in the preceding decades, and the number of mechanistically novel treatments for schizophrenia has been disappointingly low[94]. One reason for this poor performance of translational research in schizophrenia is the difficulty of applying the methods of modern drug discovery to a disorder whose pathophysiology was incompletely understood[94]. Although genetics is essential in this context, the identification of genetic variants by themselves, even if they are causative, does not mean that a viable drug target or treatment approach has been found, as the example of Huntington's disease shows. Translation and drug development in psychiatry also face other bottlenecks such as a high degree of placebo responses, tolerability problems, regulatory issues and implementation of new therapies in clinical practice, which await solution.

A functional characterization of genetic (and environmental) risk is necessary to better identify translational entry points. For this, neuroimaging alone is not enough. Cellular models, and ideally access to neuronal tissue, are necessary to understand molecular and cell architectural changes in schizophrenia, clarify epigenetic mechanisms, and to develop molecular biomarkers. There is considerable potential in integrating across cellular and systems levels to develop multivariate biomarker panels. Animal models similarly need to be improved. However, systems-level neuroscience of risk mechanisms already provides some approaches that can help translation in several ways.

First, the characterization of molecular risk mechanisms provides a quantitative entry point for computational neuroscience approaches in translation. For example, it can be quantified relatively easily to what degree a given genetic variant impacts on the abundance or activity of the gene product; however, there is currently no principled way of inferring the systems-level consequences that such a change might have. An example that this approach is feasible is the application of the theory of dopamine modulation of prefrontal cortex computational dynamics to the differential effects of the rs4680 *COMT* variant[95]. A biophysically realistic computational model, which will map the global processing features discussed above together with enough detail on genetic effects in the synapse, would be useful to define and refine our understanding of precisely how risk mechanisms affect brain function and what neural computations are most vulnerable to them. This could potentially lead to a new generation of *in silico* psychopharmacology in which the effect of a drug with a given receptor-binding profile can be linked to a predicted systems-level response.

Second, understanding neural risk mechanisms can help in personalization of existing therapy. An example is again provided by the *COMT* rs4680 variant: the reduced prefrontal efficiency associated with rs4680 Val alleles predicts that subjects carrying this variant should preferentially profit from dopaminergic stimulation. This is in fact what

has been observed for therapy with the COMT inhibitor tolcapone[96], providing a proof of principle that an understanding of the neural effects of this variant through a combined imaging genetics approach can contribute to personalized procognitive therapy. Importantly, rs4680 also predicted prefrontal activation and working memory performance under antipsychotic therapy with olanzapine[97].

Third, systems-level data can be helpful in designing a new generation of animal models. By definition, schizophrenia is a human-specific disease, because it affects human-specific faculties such as language. This does not mean that the pathophysiology of schizophrenia also needs to be human specific, but the question remains on how to optimally model relevant aspects of schizophrenia in animal models that are an essential requirement for drug discovery. By delineating neural systems properties that are consistently implicated in schizophrenia and ascertaining which behavioural features in a rodent are affected by them, a new generation of valid animal models can be designed. The molecular predictiveness of these models can be further enhanced by using genetically designed models to mimic a genetic risk variant associated with the disorder. This latter strategy will be the more promising as the amount of risk that can be attributed to the genetic risk factors increases. Therefore, modelling microdeletions could be especially fruitful. An impressive example is the recent discovery that mouse models for the schizophrenia-associated microdeletion 22q11 show abnormal hippocampal-prefrontal connectivity[98] (Fig. 4c). Defining such systems-level features through animal neuroimaging and behavioural testing and relating them to behaviour will be essential in understanding what corresponds, in a rodent, to the human-specific symptoms of schizophrenia, an endeavour that may well lead to several definable subsyndromes that will by themselves constitute a useful development for drug discovery.

Finally, not all drug development is done with animal models. In fact, a useful entry point for systems-level neuroscience could be phase 1 studies, when new substances are first introduced into humans. At this point the question arises for which mental disorders, if any, that substance might be useful; the current ability to predict efficacy is poor[94]. Here systems-level neuroscience, especially neuroimaging, may make a contribution to proof of concept at an early stage by showing whether and to what degree new substances modify the relevant systems-level features, such as disturbed connectivity or neural oscillations in healthy humans. For this, it would not even be necessary (although it would certainly be advantageous) to have these systems-level features on the causal pathway to the disorder—as the example of striatal dopamine D2 blockade in currently available antipsychotics shows, there is no intrinsic necessity for an effective therapy to intervene in the causal pathway of the disorder, and progress may be made simply by using neuroimaging endpoints rather than traditional clinical endpoints[99]. Either way, the predictive value of this approach might be even further enhanced by stratifying healthy human subjects by common genetic risk factors that are related to risk for schizophrenia, such as the SNPs discussed above, which have been shown to bias neurocircuits also implicated in manifest disorder. In addition, many features of schizophrenia psychopathology can be transiently induced in humans; for example, it is possible to produce psychotic features or cognitive dysfunction using psychotomimetic drugs. This would constitute a revival and focusing of experimental medicine in psychiatry incorporating systems-level neuroscience in early drug trials. This concept, whose analogues have been extraordinarily fruitful in oncology and haematology, now awaits application to translation in psychiatry.

1. Editorial. A decade for psychiatric disorders. *Nature* **463,** 9 (2010).
2. Weinberger, D. R. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44,** 660–669 (1987).
   A landmark conceptualization of schizophrenia as a neurodevelopmental disorder.
3. Ellison-Wright, I., Glahn, D. C., Laird, A. R., Thelen, S. M. & Bullmore, E. The anatomy of first-episode and chronic schizophrenia: an anatomical likelihood estimation meta-analysis. *Am. J. Psychiatry* **165,** 1015–1023 (2008).

4. Boos, H. B., Aleman, A., Cahn, W., Hulshoff Pol, H. & Kahn, R. S. Brain volumes in relatives of patients with schizophrenia: a meta-analysis. *Arch. Gen. Psychiatry* **64**, 297–304 (2007).

5. Goldman, A. L. *et al.* Heritability of brain morphology related to schizophrenia: a large-scale automated magnetic resonance imaging segmentation study. *Biol. Psychiatry* **63**, 475–483 (2008).

6. Lewis, D. A. & Sweet, R. A. Schizophrenia from a neural circuitry perspective: advancing toward rational pharmacological therapies. *J. Clin. Invest.* **119**, 706–716 (2009).

7. Selemon, L. D. & Goldman-Rakic, P. S. The reduced neuropil hypothesis: a circuit based model of schizophrenia. *Biol. Psychiatry* **45**, 17–25 (1999).

8. Hashimoto, T. *et al.* Gene expression deficits in a subclass of GABA neurons in the prefrontal cortex of subjects with schizophrenia. *J. Neurosci.* **23**, 6315–6326 (2003).

9. Byne, W., Hazlett, E. A., Buchsbaum, M. S. & Kemether, E. The thalamus and schizophrenia: current status of research. *Acta Neuropathol.* **117**, 347–368 (2009).

10. Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).

11. Goldman-Rakic, P. S. Working memory dysfunction in schizophrenia. *J. Neuropsychiatry Clin. Neurosci.* **6**, 348–357 (1994).

12. Callicott, J. H. *et al.* Physiological dysfunction of the dorsolateral prefrontal cortex in schizophrenia revisited. *Cereb. Cortex* **10**, 1078–1092 (2000).

13. Tan, H. Y. *et al.* Dysfunctional prefrontal regional specialization and compensation in schizophrenia. *Am. J. Psychiatry* **163**, 1969–1977 (2006).

14. Fusar-Poli, P. *et al.* Neurofunctional correlates of vulnerability to psychosis: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **31**, 465–484 (2007).

15. Achim, A. M. & Lepage, M. Episodic memory-related activation in schizophrenia: meta-analysis. *Br. J. Psychiatry* **187**, 500–509 (2005).

16. Murray, G. K. *et al.* Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol. Psychiatry* **13**, 267–276 (2008).
Evidence for abnormal salience processing in midbrain in schizophrenia.

17. Juckel, G. *et al.* Dysfunction of ventral striatal reward prediction in schizophrenia. *Neuroimage* **29**, 409–416 (2006).

18. Simon, J. J. *et al.* Neural correlates of reward processing in schizophrenia—Relationship to apathy and depression. *Schizophr. Res.* **118**, 154–161 (2009).

19. Aleman, A. & Kahn, R. S. Strange feelings: do amygdala abnormalities dysregulate the emotional brain in schizophrenia? *Prog. Neurobiol.* **77**, 283–298 (2005).

20. Rasetti, R. *et al.* Evidence that altered amygdala activity in schizophrenia is related to clinical state and not genetic risk. *Am. J. Psychiatry* **166**, 216–225 (2009).

21. Gur, R. E. *et al.* Limbic activation associated with misidentification of fearful faces and flat affect in schizophrenia. *Arch. Gen. Psychiatry* **64**, 1356–1366 (2007).

22. Brunet-Gouet, E. & Decety, J. Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. *Psychiatry Res.* **148**, 75–92 (2006).

23. Walter, H. Dysfunction of the social brain is modulated by intention type: an FMRI study. *Soc. Cogn. Affect. Neurosci.* **4**, 166–176 (2009).

24. Dierks, T. *et al.* Activation of Heschl's gyrus during auditory hallucinations. *Neuron* **22**, 615–621 (1999).

25. Hubl, D. *et al.* Pathways that make voices: white matter changes in auditory hallucinations. *Arch. Gen. Psychiatry* **61**, 658–668 (2004).

26. Wolf, R. C. *et al.* Temporally anticorrelated brain networks during working memory performance reveal aberrant prefrontal and hippocampal connectivity in patients with schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **33**, 1467–1473 (2009).

27. Whitfield-Gabrieli, S. *et al.* Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc. Natl Acad. Sci. USA* **106**, 1279–1284 (2009).

28. Meyer-Lindenberg, A. S. *et al.* Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Arch. Gen. Psychiatry* **62**, 379–386 (2005).

29. Crossley, N. A. *et al.* Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Hum. Brain Mapp.* **30**, 4129–4137 (2009).

30. Meyer-Lindenberg, A. *et al.* Evidence for abnormal cortical functional connectivity during working memory in schizophrenia. *Am. J. Psychiatry* **158**, 1809–1817 (2001).

31. Friston, K. J., Frith, C. D., Fletcher, P., Liddle, P. F. & Frackowiak, R. S. Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb. Cortex* **6**, 156–164 (1996).

32. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci.* **10**, 186–198 (2009).

33. Bassett, D. S. *et al.* Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* **28**, 9239–9248 (2008).

34. Bertolino, A. *et al.* Altered development of prefrontal neurons in rhesus monkeys with neonatal mesial temporo-limbic lesions: a proton magnetic resonance spectroscopic imaging study. *Cereb. Cortex* **7**, 740–748 (1997).

35. Braff, D. L. & Geyer, M. A. Sensorimotor gating and schizophrenia. Human and animal model studies. *Arch. Gen. Psychiatry* **47**, 181–188 (1990).

36. Jaskiw, G. E., Karoum, F. K. & Weinberger, D. R. Persistent elevations in dopamine and its metabolites in the nucleus accumbens after mild subchronic stress in rats with ibotenic acid lesions of the medial prefrontal cortex. *Brain Res.* **534**, 321–323 (1990).

37. Seeman, P. & Lee, T. Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science* **188**, 1217–1219 (1975).

38. Davis, K. L., Kahn, R. S., Ko, G. & Davidson, M. Dopamine in schizophrenia: a review and reconceptualization. *Am. J. Psychiatry* **148**, 1474–1486 (1991).

39. Laruelle, M. Imaging dopamine transmission in schizophrenia. A review and meta-analysis. *Q. J. Nucl. Med.* **42**, 211–221 (1998).

40. Howes, O. D. *et al.* Elevated striatal dopamine function linked to prodromal signs of schizophrenia. *Arch. Gen. Psychiatry* **66**, 13–20 (2009).

41. Meyer-Lindenberg, A. *et al.* Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. *Nature Neurosci.* **5**, 267–271 (2002).

42. Fusar-Poli, P. *et al.* Abnormal prefrontal activation directly related to pre-synaptic striatal dopamine dysfunction in people at clinical high risk for psychosis. *Mol. Psychiatry* doi:10.1038/mp.2009.108 (1 December 2009).
Striatal dopamine dysfunction correlates with prefrontal activation abnormalities in high-risk subjects.

43. Kapur, S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am. J. Psychiatry* **160**, 13–23 (2003).
An important conceptualization of psychosis linking it to dopamine-related salience signalling.

44. Munafo, M. R., Brown, S. M. & Hariri, A. R. Serotonin transporter (5-HTTLPR) genotype and amygdala activation: a meta-analysis. *Biol. Psychiatry* **63**, 852–857 (2008).

45. Mier, D., Kirsch, P. & Meyer-Lindenberg, A. Neural substrates of pleiotropic action of genetic variation in *COMT*: a meta-analysis. *Mol. Psychiatry* **15**, 918–927 (2010).

46. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
One of three large GWAS of schizophrenia in 2009, this paper also provides evidence for multiple common variants contributing to risk for schizophrenia that overlap with bipolar disorder.

47. Nicodemus, K. K. *et al.* Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Hum. Genet.* **127**, 441–452 (2010).

48. Liu, J. *et al.* Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum. Brain Mapp.* **30**, 241–255 (2009).

49. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* **160**, 636–645 (2003).

50. Fan, J. B. *et al.* Catechol-O-methyltransferase gene Val/Met functional polymorphism and risk of schizophrenia: a large-scale association study plus meta-analysis. *Biol. Psychiatry* **57**, 139–144 (2005).

51. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).

52. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).

53. Egan, M. F. *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl Acad. Sci. USA* **98**, 6917–6922 (2001).

54. Gogos, J. A. *et al.* Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc. Natl Acad. Sci. USA* **95**, 9991–9996 (1998).

55. Chen, J. *et al.* Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am. J. Hum. Genet.* **75**, 807–821 (2004).

56. Honea, R. *et al.* Impact of interacting functional variants in COMT on regional gray matter volume in human brain. *Neuroimage* **45**, 44–51 (2009).

57. Meyer-Lindenberg, A. *et al.* Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nature Neurosci.* **8**, 594–596 (2005).

58. Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* **71**, 877–892 (2002).

59. Barros, C. S. *et al.* Impaired maturation of dendritic spines without disorganization of cortical cell layers in mice lacking NRG1/ErbB signaling in the central nervous system. *Proc. Natl Acad. Sci. USA* **106**, 4507–4512 (2009).

60. Mei, L. & Xiong, W. C. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nature Rev. Neurosci.* **9**, 437–452 (2008).

61. Hall, J. *et al.* A neuregulin 1 variant associated with abnormal cortical function and psychotic symptoms. *Nature Neurosci.* **9**, 1477–1478 (2006).

62. Gruber, O. *et al.* Neuregulin-1 haplotype HAP(ICE) is associated with lower hippocampal volumes in schizophrenic patients and in non-affected family members. *J. Psychiatr. Res.* **43**, 1–6 (2008).

63. Mata, I. *et al.* A neuregulin 1 variant is associated with increased lateral ventricle volume in patients with first-episode schizophrenia. *Biol. Psychiatry* **65**, 535–540 (2009).

64. McIntosh, A. M. *et al.* The effects of a neuregulin 1 variant on white matter density and integrity. *Mol. Psychiatry* **13**, 1054–1059 (2008).

65. Millar, J. K. *et al.* Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* **9**, 1415–1423 (2000).

66. Ishizuka, K., Paek, M., Kamiya, A. & Sawa, A. A review of Disrupted-In-Schizophrenia-1 (DISC1): neurodevelopment, cognition, and mental conditions. *Biol. Psychiatry* **59**, 1189–1197 (2006).

67. Callicott, J. H. *et al.* Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proc. Natl Acad. Sci. USA* **102**, 8627–8632 (2005).

68. Prata, D. P. *et al.* Effect of disrupted-in-schizophrenia-1 on pre-frontal cortical function. *Mol. Psychiatry* **13**, 915–917 (2008).

69. Di Giorgio, A. *et al.* Association of the SerCys DISC1 polymorphism with human hippocampal formation gray matter and function during memory encoding. *Eur. J. Neurosci.* **28**, 2129–2136 (2008).

70. Cannon, T. D. et al. Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Arch. Gen. Psychiatry* **62**, 1205–1213 (2005).

71. Nicodemus, K. K. et al. Evidence for statistical epistasis between catechol-O-methyltransferase (COMT) and polymorphisms in RGS4, G72 (DAOA), GRM3, and DISC1: influence on risk of schizophrenia. *Hum. Genet.* **120**, 889–906 (2007).

72. Harrison, P. J. & Weinberger, D. R. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol. Psychiatry* **10**, 40–68 (2005).

73. Garcia, R. A., Vasudevan, K. & Buonanno, A. The neuregulin receptor ErbB-4 interacts with PDZ-containing proteins at neuronal synapses. *Proc. Natl Acad. Sci. USA* **97**, 3596–3601 (2000).

74. Mata, I. et al. Additive effect of NRG1 and DISC1 genes on lateral ventricle enlargement in first episode schizophrenia. *Neuroimage* **53**, 1016–1022 (2009).

75. O'Donovan, M. C. et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genet.* **40**, 1053–1055 (2008). **Identification of the *ZNF804A* variant with genome-wide support.**

76. Esslinger, C. et al. Neural mechanisms of a genome-wide supported psychosis variant. *Science* **324**, 605 (2009). **The first imaging genetics study on a genome-wide significant variant, showing effects on dorsolateral prefrontal cortex connectivity mirroring those in patients with schizophrenia.**

77. Walters, J. T. et al. Psychosis susceptibility gene *ZNF804A* and cognitive performance in schizophrenia. *Arch. Gen. Psychiatry* **67**, 692–700 (2010).

78. Lencz, T. et al. A schizophrenia risk gene, ZNF804A, influences neuroanatomical and neurocognitive phenotypes. *Neuropsychopharmacology* **35**, 2284–2291 (2010).

79. Walter, H. et al. Effects of a genome-wide supported psychosis risk variant on neural activation during a theory-of-mind task. *Mol. Psychiatry* doi:10.1038/mp.2010.18 (16 March 2010).

80. Ferreira, M. A. et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature Genet.* **40**, 1056–1058 (2008).

81. Erk, S. et al. Brain function in carriers of a genome-wide supported bipolar disorder variant. *Arch. Gen. Psychiatry* **67**, 803–811 (2010).

82. Wessa, M. et al. The *CACNA1C* risk variant for bipolar disorder influences limbic activity. *Mol. Psychiatry* doi:10.1038/mp.2009.103 (30 March 2010).

83. Ben-Shachar, S. et al. Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J. Med. Genet.* **46**, 382–388 (2009).

84. Karayiorgou, M., Simon, T. J. & Gogos, J. A. 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nature Rev. Neurosci.* **11**, 402–416 (2010).

85. Meechan, D. W., Maynard, T. M., Gopalakrishna, D., Wu, Y. & LaMantia, A. S. When half is not enough: gene expression and dosage in the 22q11 deletion syndrome. *Gene Expr.* **13**, 299–310 (2007).

86. Kempf, L. et al. Functional polymorphisms in *PRODH* are associated with risk and protection for schizophrenia and fronto-striatal structure and function. *PLoS Genet.* **4**, e1000252 (2008).

87. Lieberman, J. A., Sheitman, B. B. & Kinon, B. J. Neurochemical sensitization in the pathophysiology of schizophrenia: deficits and dysfunction in neuronal regulation and plasticity. *Neuropsychopharmacology* **17**, 205–229 (1997).

88. Selten, J. P. & Cantor-Graae, E. Social defeat: risk factor for schizophrenia? *Br. J. Psychiatry* **187**, 101–102 (2005).

89. Pruessner, J. C., Champagne, F., Meaney, M. J. & Dagher, A. Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: a positron emission tomography study using [^{11}C]raclopride. *J. Neurosci.* **24**, 2825–2831 (2004).

90. Soliman, A. et al. Stress-induced dopamine release in humans at risk of psychosis: a [^{11}C]raclopride PET study. *Neuropsychopharmacology* **33**, 2033–2041 (2008).

91. Zink, C. F. et al. Know your place: neural processing of social hierarchy in humans. *Neuron* **58**, 273–283 (2008). **An environmental stressor (unstable hierarchy) impacts on circuitry for regulation of negative affect.**

92. Caspi, A. et al. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene × environment interaction. *Biol. Psychiatry* **57**, 1117–1127 (2005).

93. Stokes, P. R. et al. Significant decreases in frontal and temporal [^{11}C]-raclopride binding after THC challenge. *Neuroimage* **52**, 1521–1527 (2010).

94. Insel, T. R. & Scolnick, E. M. Cure therapeutics and strategic prevention: raising the bar for mental health research. *Mol. Psychiatry* **11**, 11–17 (2006).

95. Durstewitz, D. & Seamans, J. K. The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biol. Psychiatry* **64**, 739–749 (2008).

96. Apud, J. A. et al. Tolcapone improves cognition and cortical information processing in normal human subjects. *Neuropsychopharmacology* **32**, 1011–1020 (2007). **Proof of principle of genotype-directed personalized therapy guided by neuroimaging mechanisms in psychiatry.**

97. Bertolino, A. et al. Interaction of COMT (Val(108/158)Met) genotype and olanzapine treatment on prefrontal cortical function in patients with schizophrenia. *Am. J. Psychiatry* **161**, 1798–1805 (2004).

98. Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A. & Gordon, J. A. Impaired hippocampal-prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature* **464**, 763–767 (2010). **A mouse model of a genetic high-risk microdeletion for schizophrenia exhibits a prefrontal-hippocampal connectivity phenotype.**

99. Tost, H. et al. Acute D2 receptor blockade induces rapid, reversible remodeling in human cortical-striatal circuits. *Nature Neurosci.* **13**, 920–922 (2010).

100. Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).

101. Sohal, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459**, 698–702 (2009).

102. von Stein, A., Chiang, C. & Konig, P. Top-down processing mediated by interareal synchronization. *Proc. Natl Acad. Sci. USA* **97**, 14748–14753 (2000).

103. Lisman, J. & Buzsaki, G. A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophr. Bull.* **34**, 974–980 (2008).

104. Sirota, A. et al. Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron* **60**, 683–697 (2008).

105. Homayoun, H. & Moghaddam, B. NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. *J. Neurosci.* **27**, 11496–11500 (2007).

106. Ito, H. T. & Schuman, E. M. Frequency-dependent gating of synaptic transmission and plasticity by dopamine. *Front Neural Circuits* **1**, 1 (2007).

107. Cho, R. Y., Konecky, R. O. & Carter, C. S. Impairments in frontal cortical gamma synchrony and cognitive control in schizophrenia. *Proc. Natl Acad. Sci. USA* **103**, 19878–19883 (2006).

108. Hong, L. E. et al. Sensory gating endophenotype based on its neural oscillatory pattern and heritability estimate. *Arch. Gen. Psychiatry* **65**, 1008–1016 (2008).

109. Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).

110. Gurden, H., Tassin, J. P. & Jay, T. M. Integrity of the mesocortical dopaminergic system is necessary for complete expression of *in vivo* hippocampal-prefrontal cortex long-term potentiation. *Neuroscience* **94**, 1019–1027 (1999).

111. Uhlhaas, P. J. et al. The development of neural synchrony reflects late maturation and restructuring of functional networks in humans. *Proc. Natl Acad. Sci. USA* **106**, 9866–9871 (2009).

112. Giorgio, A. et al. Longitudinal changes in grey and white matter during adolescence. *Neuroimage* **49**, 94–103 (2010).

The environment and schizophrenia

Jim van Os[1,2], Gunter Kenis[1] & Bart P. F. Rutten[1]

Psychotic syndromes can be understood as disorders of adaptation to social context. Although heritability is often emphasized, onset is associated with environmental factors such as early life adversity, growing up in an urban environment, minority group position and cannabis use, suggesting that exposure may have an impact on the developing 'social' brain during sensitive periods. Therefore heritability, as an index of genetic influence, may be of limited explanatory power unless viewed in the context of interaction with social effects. Longitudinal research is needed to uncover gene–environment interplay that determines how expression of vulnerability in the general population may give rise to more severe psychopathology.

The disorder schizophrenia, diagnosed in around 0.5–1.0% of the population during their lifetimes[1], may be considered the poor outcome fraction of a truly 'complex', multidimensional psychotic syndrome (lifetime prevalence 2–3%, ref. 2), that in turn can be traced to measurable, age-dependent (young > old) expression of liability in a substantial proportion—around 10–20%, ref. 3—of the non-ill general population[4]. The correlated symptom dimensions of the psychotic syndrome are: psychosis (hallucinations and delusions), motivational impairment (avolition or amotivation), affective dysregulation (depression, mania) and alterations in information processing (cognitive impairment).

High heritability estimates indicate a strong genetic influence. Although the well known 'stress-vulnerability' model of aetiological influence in psychiatry assumes that genetic factors operate by making individuals selectively vulnerable for environmental risks (gene–environment interaction, or GxE), it has proven very difficult to provide data substantiating this supposition. However, recent evidence of substantial variation in the incidence across places and minority groups, associated with high attributable fractions[1,5–8], suggests that environmental factors do have an important role. Here, we will review the evidence linking environmental risks to psychotic syndrome, and examine to what degree such associations may be methodologically valid and indicative of causal influence; which aspects of the environment actually explain, or mediate (as factor between the environmental exposure and the outcome that actually occasions the effect), for example, variation in incidence across places; and which biological and cognitive mechanisms may underlie such effects. Finally, is there evidence that genetic influence for psychotic syndrome may operate in part by creating subgroups that are more vulnerable to environmental risks and, if so, how should the issue of genetically influenced sensitivity to the environment be developed further in clinical and translational research?

Phenotype, environment and heritability

Whereas classical twin studies (the study of twins without inclusion of twins' relatives) indicate that shared environmental effects are small, quantitative biometric analyses show that this type of twin methodology represents a singularly poor method to characterize the effects of the environment in any way—including the argument that shared environmental effects are small[9]. Furthermore, there is evidence that the level of familial clustering of psychotic disorder is greater if it is measured in risk environments, for example urban environment or minority group[10,11], suggesting that heritability estimates from classical twin studies not only reflect genetic influence, but also underlying gene–environment interactions[12,13]. Thus, the heritability estimates for the behavioural expressions of liability for psychotic disorder in the non-patient general population, indexed by subtle manifestation of psychotic experiences[14], affective dysregulation[15], motivational impairments[16] and cognitive alterations[17], are in the order of 40–70%. Interaction between the genes influencing these liability traits and environmental risk factors may give rise to more 'co-morbid'[18] disease phenotypes with higher heritability estimates (as they now also include GxE effects), and greater probability of passing the 'filters' on the pathway to mental health services[19,20] (Fig. 1).

Cities, minorities, trauma and drug use

Increasingly sophisticated studies, summarized in a growing body of meta-analytical work, suggest that psychotic outcomes are associated with growing up in an urbanized area, minority group position, cannabis use and developmental trauma. Relative risks are mostly in the order of 2, although as high as 5 in certain subgroups. However, definitive conclusions about association between environment and psychotic syndrome critically depend on a number of qualitative and quantitative issues described in Box 1.

The evidence for developmental trauma

Although systematic reviews on the association between developmental trauma and psychosis are not consistent[21–23], an explosion of new studies have since consistently demonstrated dose–response associations across a range of designs, natural experiments and endpoints, including a number of strong prospective studies establishing temporal order and ruling out reverse causality[24–26]. Reports of developmental trauma in patients may be different compared to controls due to the presence of psychotic symptoms or because of patients searching for reasons for their predicament; studies have therefore introduced (semi-) prospective designs[24,26], assessment of trauma through independent sources[25,26] and validation procedures for reporting of trauma by patients[27]. Studies have addressed genetic confounding by controlling, directly or indirectly, for genetic risk[24,25,28]. The association with developmental trauma may be mediated by a combination of both neglect and abuse[21–23]; however, more work is needed to examine the important issue of what mediates the effect.

The evidence for minority group position

Meta-analytical work shows consistency for the association between psychotic syndrome and minority group position across a wide range

[1]European Graduate School for Neuroscience, SEARCH, Maastricht University Medical Centre, 6200 MD Maastricht, The Netherlands. [2]King's College London, King's Health Partners, Department of Psychosis Studies, Institute of Psychiatry, London SE5 8AF, UK.
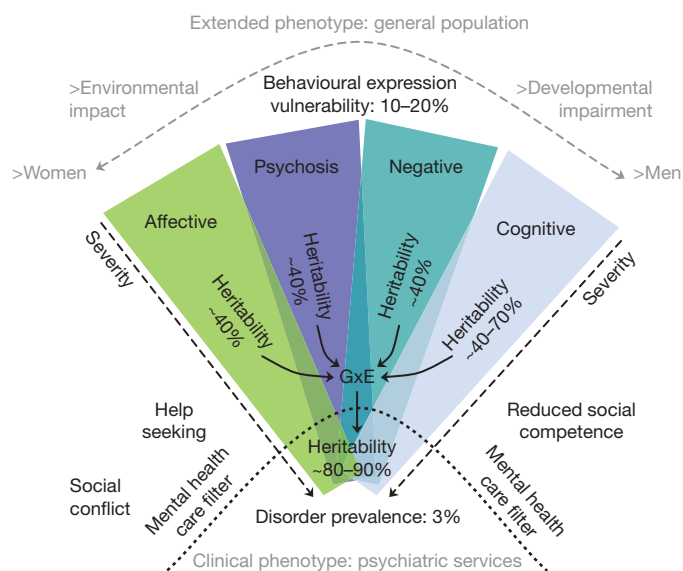
Figure 1 | Complexity of the psychotic disorder phenotype in aetiological research. Four main dimensions of affective dysregulation (depression, mania, anxiety), psychosis (delusions, hallucinations), negative symptoms (motivational impairment) and cognitive alterations characterize the general psychotic syndrome. In the general population, low levels of these dimensional phenotypes represent the behavioural expression of vulnerability (extended phenotype in figure, prevalence 10–20%), which display moderately high heritability levels, and display low levels of correlation (as indicated by less overlap of the four dimensions at the level of behavioural expression of vulnerability). Gene–environment interactions give rise to more severe levels of phenotypic expression, which in turn contribute to likelihood of clinical detection by psychiatric services by passing the filters on the pathway to mental health care (mental health care filter in figure) in a dimension-specific fashion. For example, higher levels of affective dysregulation are likely to lead to active help-seeking, whereas psychosis may lead to clinical detection through social conflict and cognitive symptoms through reduced social competence. As dimensions contribute independently to clinical detection by psychiatric services, symptom dimensions are subject to co-morbidity bias (that is, are much more correlated and 'co-morbid' at the level of psychiatric services), giving rise to clinical diagnostic descriptions such as schizophrenia that have a relatively low prevalence. These diagnostic descriptions have higher heritability as they include the effect of gene–environment interactions underlying severity of dimensional expression. Evidence suggests that environmental influences associated with psychotic disorder may have greater impact on affective and psychotic dimensions, which are more common in women, whereas negative and cognitive dimensions may be more strongly associated with developmental impairment and male sex. GxE, gene–environment interaction.

of approaches, endpoints, settings and cultural group definitions[5,29], and after adjustment for a range of confounders. The possibility of cultural bias in diagnosis and selective migration has been examined but not found to have a major impact on the association. The association with minority group position is observed in both first and second generation migrants[5,29], as well as in minority groups without recent migration[30], indicating that pre-migration factors or migration itself are unlikely to mediate effects. Studies in four different countries have shown that the effect of minority ethnic group on psychotic syndrome depends on the ethnic density of the area the person is living in: the greater the proportion of the own ethnic group in the area, the lower the risk for psychotic disorder[31,32]. These findings suggest that it is not ethnic group per se that increases risk, but rather the degree to which one occupies a minority position, or stands out in relation to the wider social environment. Additional research suggests that effects associated with minority group position may be mediated by chronic social adversity and discrimination[33], resulting in social marginalization or a state of social 'defeat' (chronic experience of an inferior position or social exclusion[34]).

BOX 1

Weighing the epidemiological evidence for environmental association and causality

• Methodological explanations, such as bias or (genetic) confounding, must be ruled out.

Bias may arise, for example, when cases or controls are selected, followed over time, or interviewed in such a way that they are more or less likely to report environmental exposure in the case-control comparison (for example, a case-control study of psychotic syndrome and cannabis use recruiting controls from a selected community with very strict lifestyle rules).

Confounding occurs when, for example, an apparent difference in exposure rate between cases and controls is due to a third factor associated with both exposure and, independent of that, with the illness (for example, cases are much younger than controls therefore report more cannabis use resulting in a spurious association between cannabis and disease).

Genetic confounding refers to the fact that, for example, genetic liability for schizophrenia may predispose to cannabis use, so that the association between cannabis use and schizophrenia in fact represents a genetic epiphenomenon. Similarly, genes predisposing for psychotic syndrome may cause selective migration to another country, resulting in minority group position, or an urban environment. Genetic confounding may be unlikely a priori, as the genetic effect on cannabis use would need to be very substantial. However, a degree of genetic confounding is difficult to exclude.

• It is critical that observational research provides consistent evidence for an association between illness and environmental risk across a mix of different strategies, natural experiments, designs, endpoints and levels of genetic sensitivity. If this is the case, validity is more likely.

• Exposures such as urban environment and minority position really represent proxies for as yet unidentified environmental factors, which represent the mediators of the effect, that require identification.

In humans, experimental support for causality, in the sense of randomized allocation to exposure, is only possible for acute outcomes such as induction of transient delusions or hallucinations or cognitive impairment following experimental cannabis use, experimental stress or other experimental situations. Causality from observational studies may be inferred using criteria of consistency of the association, temporal order (that is, the exposure precedes the outcome and reverse causality is ruled out, for example early psychotic symptoms causing cannabis use rather than the other way round), evidence for dose–response (more exposure results in progressively greater risk) and a plausible link to underlying biological and cognitive mechanisms.

The evidence for growing up in an urban environment

Meta-analytical work shows consistency in a dose–response association with the urban environment across a wide range of settings, endpoints, data collection approaches and definitions of 'urbanicity'[1,6–8]. Confounding can occur as, for example, high rates of schizophrenia in cities could be secondary to higher rates of drug use or ethnic minority groups living in urban environments; however, studies have controlled for a wide range of possible confounders[6–8], including variables indexing genetic risk in order to exclude genetic confounding[6–8,35]. Longitudinal studies of natural experiments show that changing the environmental exposure, for example moving from an urban to a rural environment in childhood, brings about a corresponding decrease in risk for psychotic outcome[36], which argues against urbanicity representing a non-causal genetic epiphenomenon.

Similar to the effect of minority group position, contextual effects involving the wider social environment may point to what mediates the effect of urban environments. Thus, there is evidence that risk for psychotic syndrome associated with indicators of social maladjustment, for example single parent family, single marital status and residential instability, similarly varies with the degree to which this represents the exception in relation to the wider social environment[37,38]. This type of interaction between individual-level and area-level social 'fragmentation' may mediate the effect of the urban environment[38].

The evidence for cannabis use

Randomized experimental studies show that delta-9-tetrahydrocannabinol, the main psychotropic component of cannabis, causes transitory psychotic symptoms and impaired cognition in healthy volunteers[39], and that individuals at genetic risk for psychotic syndrome display an exaggerated psychotic response[40]. Meta-analytical work shows that the association between psychotic syndrome and cannabis is consistent, also after adjustment for a range of confounding factors[41]. Cannabis use may reflect self-medication for early expression of psychotic vulnerability or symptoms, and there is evidence that both self-medication (psychosis proneness may induce cannabis use) and causation (cannabis induces psychosis proneness) apply[42], although other work has failed to produce evidence for self-medication (ref. 43 and R. Kuepper, J. van Os, R. Lieb, H. U. Wittchen and C. Henquet, results not shown). Studies addressing genetic confounding by examining if indices of genetic risk predict exposure[44], and by controlling for genetic risk[45,46], do not suggest that genetic confounding can explain much of the observed association between cannabis and psychotic syndrome.

The association between cannabis use and psychotic outcomes has been demonstrated at the level of (1) cognition outcomes, cerebral neuroimaging phenotypes, and at the clinical level of psychotic syndrome as well as at the subclinical level of behavioural expression of vulnerability in the form of subtle psychotic experiences in non-patients; (2) in case-control, case-sibling, longitudinal (birth) cohort and cross-sectional studies; (3) in experimental studies as well as observational studies; (4) across individuals at average (healthy controls), higher than average (for example, siblings of patients) and high genetic risk (for example, patients) for psychotic syndrome and across natural variation in potency[47].

Prenatal environmental influences

A very wide variety of specific prenatal environmental exposures (for example, prenatal maternal stress, prenatal maternal nutritional deficiency, maternal serum lead and homocysteine levels, rhesus incompatibility, low and high neonatal vitamin D, prenatal toxoplasmosis, specific viral and bacterial infections, miscellaneous pregnancy and birth complications (PBC)) has been reported in relation to adult psychotic outcomes. There are, however, few true (that is, corresponding in trimester timing, exposure definition, subgroup-only effects, specific exposures within the miscellaneous group of PBC that are thought to signal hypoxia-related events during pregnancy[48]) replications, such that no definitive conclusions about association can be drawn at this stage.

Is there evidence for vulnerable subgroups?

Exposure to urban environment, cannabis use, developmental trauma and minority group position represents, in combination, a common occurrence, whereas the rate of psychotic syndrome is low. This suggests that beneath the relatively small marginal risks linking the environment to psychotic syndrome at the population level, vulnerable subgroups exist that are more sensitive to a particular environmental risk factor at a much larger effect size. Thus, the validity of observed associations with urban environment, developmental trauma, cannabis use and minority group position hinges on evidence of vulnerable subgroups. Genetically sensitive studies indicate that differential sensitivity to the psychosis-inducing effects of environmental factors may be mediated by genetic factors[12]. For example, in siblings of patients with a psychotic

disorder, who are at increased genetic risk to develop psychotic disorder, the psychotomimetic effect of cannabis is much greater than in controls[45], as is the risk to develop psychotic disorder when growing up in an urban environment[11,49] (Fig. 2).

Cognitive mechanisms of environmental impact

A large body of literature indicates that early social, cognitive and emotional development is important to later health and material outcomes. However, the cognitive alterations in schizophrenia and related psychotic disorders not only include the neuropsychological domains of attention, memory, processing speed and reasoning, but also the correlated, although not entirely overlapping, higher order domain of social cognition. Social cognition revolves around concepts such as attribution, intention, agency and emotion, that underlie the mental operations guiding social behaviour (for example, the correct interpretation of another person's intentions or emotions, referred to as mentalizing ability or 'theory of mind'[50]) and self-representation (the differentiation between the 'self' and 'other' that prevents misattribution of agency— failure to recognize oneself as the source of one's actions, thoughts or feelings, due to malfunction in the normal capacity to ignore self-generated sensations because they are predictable). In other words, social cognition is an important mediator in shaping a representation of oneself in relation to the social environment; impaired social cognition may result in aberrant representations and psychotic symptoms.

A direct link between mentalizing ability and symptoms is suggested by the fact that delusions observed in psychotic disorder frequently present as alterations in social inference; for example, paranoid delusions involve erroneous attribution of harmful intentions to behaviour observed in others. Mentalizing ability is also associated with reasoning biases observed in psychotic disorder, particularly a data gathering bias, referred to as 'jumping to conclusions', underlying severity of delusional
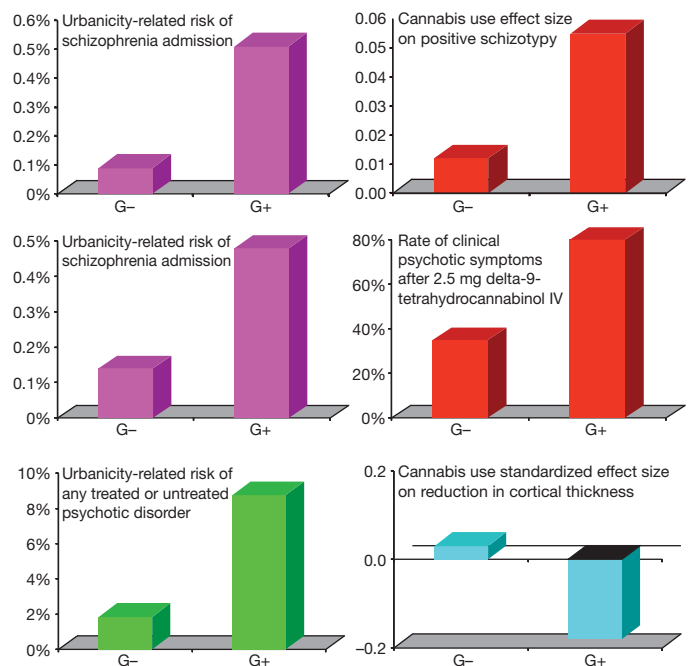
Figure 2 | Evidence for vulnerable subgroups and gene–environment interaction using proxy measures of genetic risk. Selected studies demonstrating differential sensitivity for urbanicity (left side, from top to bottom: refs 89, 49 and 11) and cannabis use (right side, from top to bottom: refs 45, 40 and 90) are shown. Colours correspond to type of psychotic outcome assessed. G+ and G− refer to, respectively, groups with and without genetic risk for psychotic disorder as defined by combination of premorbid cognition and social functioning[89], family history of psychotic disorder in first-degree relative[11,49], being a sibling of a patient with psychotic disorder[45] and being a patient with psychotic disorder[40,90].

ideation. Similarly, psychotic symptoms such as 'made' feelings or movements (passivity phenomena), thought insertion, or auditory hallucinations can be understood as impaired sense of agency, secondary to sensory prediction deficits[51], resulting in attribution of one's own actions, feelings, thoughts and inner speech to external sources. Latent vulnerability in this domain can become expressed following environmental exposure. For example, exposure to environmental variation in the form of short-term sensory deprivation or random noise has been shown to result in hallucinatory experiences in susceptible individuals[52,53].

Mentalizing ability normally emerges during preschool years and is environment-driven in that it is highly dependent on day-to-day social interactions. Selective deprivation of access to early social interactions and actively seeking out social information by, for example, hearing impairment or exposure to adversity/deprivation during critical developmental phases, may interfere with the acquisition of mentalizing ability and increase risk of later psychotic symptoms. Evidence to support this notion comes from studies showing significant delays in the mastery of mentalizing ability in hard-of-hearing[54] and maltreated[55] children on the one hand, and, on the other, an association between hearing impairment[56,57] and developmental trauma[21] and later psychotic symptoms or psychotic disorder in young people. There is evidence that other exposures that may increase risk for schizophrenia, such as head injury[58] and methamphetamine use[59], also interfere with the development of mentalizing ability[60,61], suggesting a more general link between the environment, social cognition and psychotic disorder.

The clinical relevance of social cognition is apparent in its association with course and outcome of schizophrenia and related psychotic disorders, particularly as regards social competence and community functioning. A recent study presenting 48 independent meta-analyses on associations between 12 a priori identified neuropsychological and social cognitive domains and 4 domains of functional outcome showed that social cognition was more strongly associated with community functioning than neuropsychological domains of cognition. This finding was mostly due to stronger associations with theory of mind[62], suggesting that environment-cognition-symptoms-outcome relationships in psychotic disorder may be mediated to a degree by the ability to correctly infer the mental states of others in the social environment.

Biological mechanisms of environmental impact

Exposures to environmental variation during developmentally sensitive periods are essential for the normal development of neuronal connectivity underlying functional abilities of the human brain. Early neglect and life course environmental insults that disinhibit stress signalling pathways can lead to impaired neuronal responsiveness and symptoms of profound prefrontal cortical dysfunction, providing a direct link between the environment and the cognitive impairments observed in psychotic syndrome[63].

Whereas genetic factors moderate the sensitivity of specific types of neural cells or circuits as well as the timing of environmental sensitivity during development[64], other mechanisms have been described that mediate imprinting of environment and experience, acting in parallel at different biological levels (Fig. 3b). The timing of environmental exposures associated with psychotic disorder, viewed in relation to the developmental biology of normal experience-dependent brain development (Fig. 4), is compatible with extensive developmental alterations having an impact in the areas of cognition and emotion, as observed in psychotic disorder.

Age-dependent, developmental expression of subclinical psychotic experiences in adolescence are mostly transient; however, repeated exposure to environmental risk factors may cause subclinical psychotic experiences to persist and become more severe, resulting in onset of psychotic illness in a minority of individuals[20]. These data are suggestive of a mechanism of sensitization. There is evidence that exposure to
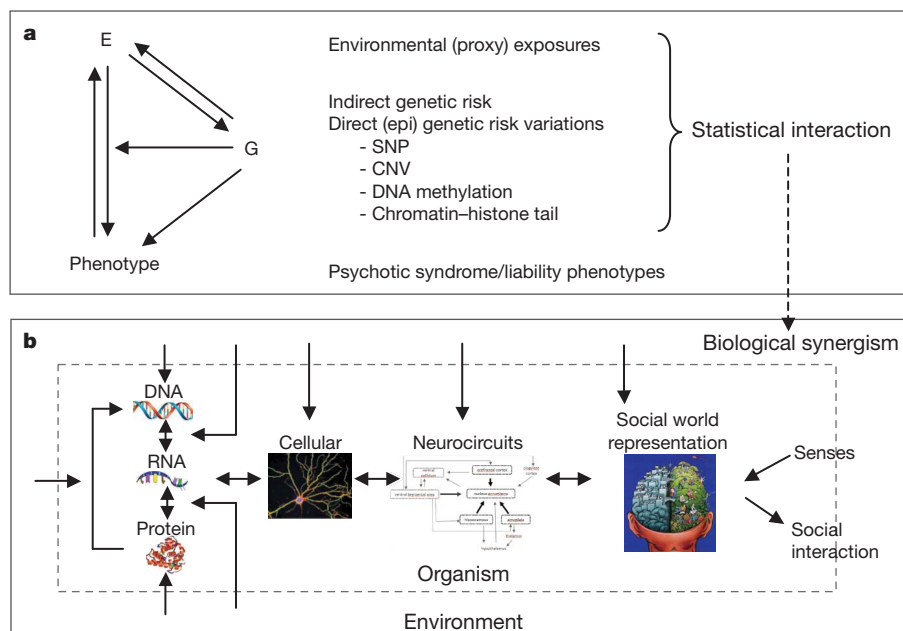
Figure 3 | Schematic illustration of gene–environment interplay at the levels of epidemiology and biology. The left part of **a** represents relationships between environmental exposures (E), genetic liability (G) and the phenotype, that is, psychotic syndrome and related liability phenotypes. Genes may control environmental exposure (arrow from G to E), genes may control environmental sensitivity (arrow from G to arrow E to phenotype), the phenotype may induce reverse causality (arrow from phenotype to E) and the environment may occasion (epigenetic) mutations (arrow from E to G). Measurement of environmental exposures often represent proxies of unknown mediating factors, and measures of genetic risk may be epidemiological (indirect) or molecular (direct). Current genetic and epigenetic variations comprise single nucleotide polymorphisms (SNP), copy number variations (CNV), DNA methylation, chromatin structure and histone tail alterations. Epidemiological studies may focus on the identification of statistically significant marginal effects of individual factors or the statistical interaction between (proxy) environmental exposures and genetic variant(s); evidence for statistical interaction may be pursued to identify biological synergism, or the co-action of genes and environment at the biological level. Panel **b** illustrates the interrelations between the various levels of biology mediating mental representation of the social world, and the numerous ways that environmental factors may impact on or interact with these. These interactions may be studied in different organisms and model systems using observational and experimental study designs.

Figure 4 | Schematic illustration of the approximate timing of the development of the human brain, functional abilities, and impact of environmental exposures. Arrows reflect impact of environmental factors associated with psychotic syndrome. Grey bars indicate the approximate developmental periods during which the processes depicted in the column are active (in **a**, around environmental exposures) or are being established/develop (in **b**, **c** and **d**). The grey colour intensities of the bars in **a**–**d**, and around the environmental exposures, indicate the approximate magnitude of the process or the approximate strength of development or maximum exposure over time. The developmental windows for the functional abilities in neurocognition, affect and social cognition (in **b**, **c** and **d**) are only given for the postnatal period. The red box indicates the window of maximum additive environmental impact.

adversity early in life renders individuals more sensitive to the effects of stress in adulthood and more prone to experience anomalous perceptions (such as flashbacks)—referred to as behavioural sensitization[65,66]. It has been proposed that early/repeated exposure to environmental risks results in increased mesolimbic dopamine reactivity[67]—known as endogenous sensitization. Exposure to many relevant environmental risk factors in animals, for example prenatal infection, prenatal stress, prenatal malnutrition, early life adversity, adolescent cannabis use, repeated psychostimulant exposure and social defeat stress, have all been shown to induce altered dopamine neurotransmission and sensitization of mesolimbic dopamine neurons in early adulthood, resulting in augmented expression of psychosis-related phenotypes. In humans, a sensitized state of the striatal dopamine system may be brought about by administration of a few doses of the psychostimulant amphetamine, another exposure associated with psychotic illness[59]. Evidence from

studies in a variety of mammalian species has shown that repeated exposure has sustained impact on regulation of mesolimbic dopaminergic neurotransmission at different biological levels: (1) molecular biological alterations including induction of transcription factors and altered chromatin plasticity; (2) chemical alterations including abnormal dopaminergic drive associated with alterations in phasic and tonic dopaminergic firing; (3) induction of several signal transduction pathways; (4) altered ratio of dopamine D2 and D1 receptor levels; (5) electrophysiological alterations; (6) structural alterations of dendritic spines; and (7) increased levels of dopamine receptors in the high-affinity state. Animal research shows that many environmental exposures (for example prenatal stress[68]) also disrupt prefrontal cortex function and corticolimbic interactions, which may be linked to observations of impaired social cognition following childhood trauma in humans. Although it may be attractive to propose that perturbation of corticolimbic circuits precedes and/or

mediates mesolimbic sensitization, research regarding this issue remains inconclusive.

Most of the work on the neurobiology of psychotic syndrome to date has focused on alterations in neurotransmitter systems. The dopamine hypothesis has survived several decades, its current version[69] postulating that multiple environmental and genetic risks during development interact to funnel through a single final common pathway of presynaptic striatal hyperdopaminergia causing delusions and hallucinations. Given that dopaminergic function is intertwined with, and regulated by, GABAergic, glutamatergic and endocannabinoid signalling, current views about environmental impact not only incorporate neurodevelopmental aberrations in these systems, but also attempt to take into account possible alterations in myelination, synapse formation, the immune system and mitochondrial energy metabolism. Figure 5 illustrates neural circuits thought to be relevant in mesolimbic neurotransmission and psychotic disorder, the impact of environmental exposures on these regions and circuits as shown in human studies, and the effect of repeated environmental exposures on top-down cognitive control over bottom-up sensory input in the nucleus accumbens, giving rise to aberrant development of mentalizing ability and self-representation. Whether or not psychotic disorder results from a single pathway of functional interaction between prefrontal network dysregulation and altered mesolimbic dopamine signalling remains to be established.

The challenge of translational approaches

Novel translational approaches using both observational and experimental human and animal studies will be essential to decipher further the consequences of exposures during environmentally sensitive periods on different biological levels during brain development as well as on psychological and behavioural phenotypes. Extension of multidisciplinary translational studies modelling environmental impacts, genetic moderation thereof, and their impact on behaviours associated with human mental illness liability phenotypes, for example social cognition, is therefore necessary. Table 1 summarizes potentially fruitful strategies, and challenges, of animal studies in this area. These strategies will be particularly fruitful when they entail the integration of (1) biological underpinning of social phenotypes, (2) biological read-outs at various cellular and molecular levels, (3) biological read-outs from various brain regions and cell types, (4) experimental designs of (repeated) environmental exposures during developmentally sensitive time windows, and (5) dynamic genetic manipulations during development, followed by multidomain behavioural phenotyping.

Mechanisms of gene–environment interplay

Given evidence that genes may have an impact on risk for psychotic syndrome by altering environmental sensitivity, gene–environment interaction research is a logical next step but in practice remains very rare, because to date there has been little collaboration between 'environmental' and 'genetic' groups in this area. Gene–environment interaction studies are extremely cost-effective, as both genetic and environmental information is collected in a single effort under the same phenotypic assessment, replacing the current duplicated and unrelated efforts to collect genetic information in one sample and environmental information in another.
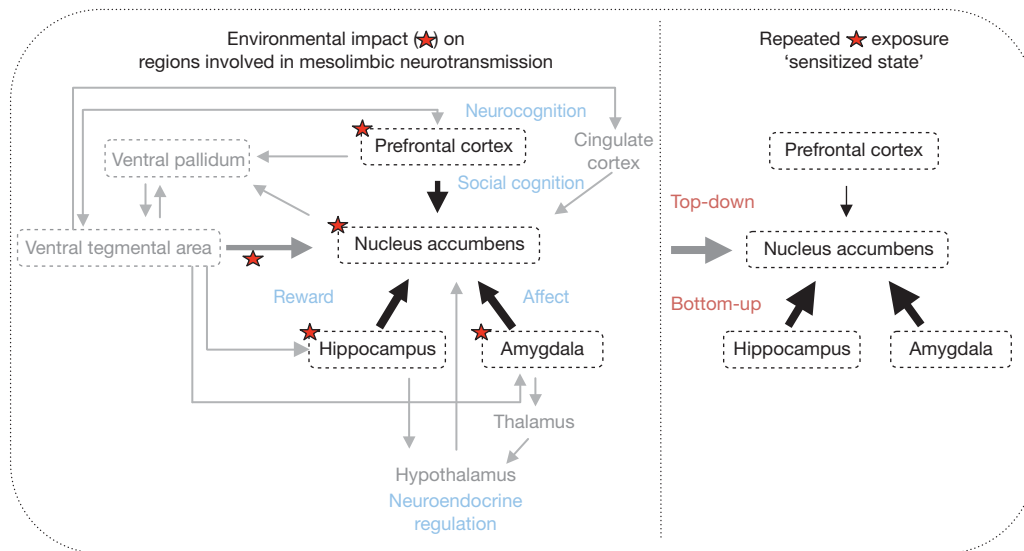
Figure 5 | Schematic overview of neural circuits thought to be relevant in mesocorticolimbic neurotransmission and psychotic disorder, illustration of reported environmental impact on these regions and circuits in human studies, and the effect of repeated environmental exposures on the regulation of bottom-up sensory input and top-down cognitive control by the nucleus accumbens. The left panel schematically summarizes the main brain regions and their interconnections involved in the pathophysiology of schizophrenia. Evidence suggests a close connection between positive symptoms in psychotic disorder, altered mesolimbic dopaminergic neurotransmission in the brain, and environmental exposures. Mesolimbic neurotransmission encompasses the nucleus accumbens as the site of integration of bottom-up sensory experiences with top-down cognitive control of dopamine neurotransmission, thus regulating dopamine firing in the ventral tegmental area via GABAergic feedback through the ventral pallidum. The activity of the mesocorticolimbic circuit is thus subserved by reciprocal connections with limbic and frontal cortical areas; it mediates various functions among which to maintain an updated representation of the social world. The integration is co-determined by the magnitude of the dopaminergic input from the ventral tegmental area, increased dopamine levels enhancing limbic signalling and attenuating cortical control[91]. A mesolimbic hyperdopaminergic state may thus diminish the impact of regulatory cognitive processes and contribute to aberrant representations of salient stimuli and altered reward-related learning, two features of impaired information processing in patients with psychotic syndrome. Red star symbols indicate reported associations between exposure to childhood trauma[90,92–95] and cannabis use[96–99] on the one hand, and functional or structural alterations from human studies on the other; findings of environmental impact only refer to brain regions enclosed by dashed lines. The right panel illustrates how repeated exposure to environmental factors may result in a 'sensitized state', representing a dysregulated balance between, on the one hand, activation of the nucleus accumbens by hippocampal and amygdalar projections and, on the other, inhibitory control of the nucleus accumbens by glutamatergic projections from the prefrontal cortex[91]. The resulting enhancement of bottom-up sensory experiences without appropriate top-down cognitive control is proposed to underlie aberrant representations of the social world. Note: for reasons of clarity, this schematic overview does not depict all anatomical brain regions and their interconnections.

Table 1 | Future perspectives and challenges for animal studies of psychotic disorder phenotypes

Research domain	Strategy	Challenges
Development	Identification of environmentally sensitive periods during development of higher cerebral functional abilities, and key regulatory signalling pathways.	Valid measurement of relevant functional abilities across species. Differential developmental timing between species.
Timing, severity, duration and frequency of environmental exposures in effect on adult phenotypes	Well-controlled experimental administration and manipulation of environmental exposures. Deciphering pre- and postnatal effects using cross-fostering experiments. Investigation of multiple offspring generations after environmental exposure of first generation. Combined exposure to different environmental factors.	Exact factors underlying 'proxy' risk factors currently unknown. Transferability of (experience of) social stress between humans and other animal species unclear. Temporal sequence of exposures in psychotic disorder not clear.
Genetic control of environmentally sensitive developmental windows	Dynamic manipulation of gene expression during development, for example, using Cre-lox technology and/or tetracycline-activated systems, small interference RNAs, or optogenetic manipulation of neuronal activity.	Translating neurodevelopmental periods across species. Differentiating the effects of the homologous endogenous gene that is expressed in conjunction with the corresponding human transgene. Extrapolation of rodent genetic manipulation to human genetic risk variants, including copy number variations. Deciphering putative differences in genetic drive of phenotypes in various animal species.
Multidomain phenotyping	Combined assessment of various behavioural domains: neurocognition, motivation, stress reactivity, social cognition, socially transferable learning. Combined assessment of experience-dependent adaptation at different biological levels: molecular (gene expression, epigenetics), cellular (dendritic spines), and functional (electrophysiology) levels.	Enhancing uniformity in behavioural testing across laboratories. Assessing subjective experience of psychiatric phenotypes in species other than humans.
Temporal sequence of biological alterations underlying the path from environmental exposure to psychosis domains	Studying environmental impact on salience attribution using behavioural measures of reward representation and associative learning. Linking sensitization of mesolimbic to other neurotransmitter systems, neuronal circuits, including their myelinated connections, and neuroendocrine regulation. Explore other possible mechanisms not yet clearly linked to environment and need to conduct experimental studies of environmental impact on these.	Detection of 'real' hallucinatory experiences and delusions in animals. Differentiation of mediating and moderating role of biological processes that are changed after environmental exposure.
Comparative ethology	Explore various animal species and strains for naturally occurring differences in psychosis-related phenotypes, such as social cognitive abilities, and underlying neurobiology.	Experimental manipulation of social environmental factors in different social contexts, including the natural habitat.

To date, the vast majority of gene–environment interaction studies of the psychosis syndrome have used indirect measures of genetic risk, such as genetic risk for psychotic syndrome that is shared with an ill first-degree relative, or expressed as psychotic syndrome-related cognitive impairments[12,13]. Studies modelling genetic risk as a single variable reflecting the hypothesized net genetic effect will continue to be of importance; the next wave of studies may benefit from enhanced integration of the timing, severity and experience of environmental exposures, richer genetic contrasts (for example, by including first-, second- and third-degree relatives or using extended twin-family designs), (semi-) experimental designs and more extended characterization of the relevant liability phenotypes within the same study design.

A considerable novel challenge will be to examine differential sensitivity for the environment using molecular genetic measures of risk (Fig. 3). Several approaches can be envisaged. First, interactions between molecular genetic variation (common and rare single nucleotide polymorphisms, copy number variants, epigenetic modifications) and environmental measures can be modelled using statistical approaches (Fig. 3a). Although this requires new approaches, the field is developing very rapidly, providing a range of possible solutions[70]. Even though statistical models do not provide direct information on the parameter of interest—how genes and environment jointly have an impact on biology (biological synergism)—methods exist that allow for an estimation of the relevance of the statistical interaction in relation to underlying biological synergism[71]. Statistical interaction can be examined using a hypothesis-based approach, based on insights into the biological mechanisms associated with environmental risks and molecular genetic effects, allowing for the formulation of functionally plausible hypotheses[72,73], a discovery approach using mass marker information[70], or a combination of the two[46]. Second, synergism between genes and environment can be examined dynamically, in animals or humans, at the level of an experimental biological, cerebral or psychometric phenotype (Fig. 3b). Experimental studies thus represent an important source of follow-up validation of findings from initial studies focusing on statistical interaction.

As individual genes interact in complex manners with other genes and with non-coding sequences, it may be furthermore productive to (1) combine pathway analyses in genome-wide-association studies with environmental-wide assessments and explore gene–environment-wide interactions (GEWIS)[70], (2) study the effect of complete patient genomes in interaction with specific environmental exposures by sampling accessible tissue sources and expose cultured cells to specific environmental factors to observe subsequent biological alterations, (3) study prospectively sampled biological tissues in order to investigate longitudinal changes in (regulation of) gene expression as a result of environmental exposures[74] during the follow-up period, and correlate these to longitudinal alterations in liability phenotypes. Finally, (4) recent work suggests that rare structural DNA variants, or copy number variants, contribute broadly to neurodevelopmental phenotypes including both schizophrenia and autism[4], and similar evidence of non-specific contribution to both schizophrenia and autism has been reported for minority group position[75,76] and advanced parental, particularly paternal, age[76,77]. These findings may point to specific forms of gene–environment interplay having an impact on neurodevelopmental alterations. For example, advanced paternal age may reflect a mechanism of cumulative environmental exposure affecting the male germline through epigenetic mechanisms, causing developmental perturbations that increase risk across a spectrum of neurodevelopmental disorders.

Beyond GxE: environmental sensitivity outcomes

Aetiological research in psychiatry overwhelmingly focuses on proposed disorder-specific causes. However, a diagnosis of schizophrenia is highly predictive of virtually all other Axis I and Axis II psychiatric disorders in the same person[78]. In addition, research has shown that the cognitive alterations, psychosis (hallucinations and delusions) and affective dysregulation observed in psychotic disorders (schizophrenia, bipolar disorder) are also prevalent in common mental disorders (anxiety disorders, depression), and that differences between the two groups are quantitative rather than qualitative[79,80]. The relative non-specificity of

symptoms in psychiatric disorders extends to the level of familial aggregation. For example, the non-affected siblings of patients with psychotic disorder display cognitive alterations compared to well controls, one of the reasons why cognitive alterations are considered a 'core' marker of genetic risk for schizophrenia. However, siblings of patients with common mental disorders also display cognitive alterations, albeit to a lesser degree[81]. Similarly, although a family history of schizophrenia is associated with the strongest relative risk, almost any psychiatric disorder in first-degree relatives is associated with an increased risk for schizophrenia. Indeed, in terms of attributable risk, nearly 30% of schizophrenia in the population can be attributed to psychiatric family history in general, compared to 6% that is attributable to a family history of schizophrenia specifically[35]. These data suggest that, in addition to possible specific factors impacting on liability to specific psychiatric disorders, shared genetic and environmental factors occasion neurodevelopmental alterations resulting in liability to broad dimensions of mental ill-health[82], expressed in the form of variable combinations of affective dysregulation, psychosis, motivational impairment and cognitive alterations.

For example, one of the factors that may broadly increase risk for both psychotic[83] and common mental disorders[84] is the personality trait of negative affectivity, or the stable tendency to develop negative emotions in the face of small daily life stressors. Environmental negative reactivity is difficult to capture in cross-sectional questionnaires, but can be assessed in the flow of daily life using context-sensitive momentary assessment technology, the mental health equivalent of ambulatory, 'real life' measurement of physiological parameters such as blood pressure, heart rate, brain waves and muscle tone. By sampling affective and environmental experience at random moments in the flow of daily life over protracted periods of time (typically a week or more), in combination with simultaneous, non-invasive sampling of a range of biological parameters, individual differences in affective and associated biological responses to small variations in environmental stress can be analysed directly, allowing for a comparison of different psychiatric disorders and determination of the genetic and environmental factors that contribute to individual differences in the 'film' rather than the 'snapshot' of environmental reactivity[85]. Momentary assessment research suggests that altered negative affective reactivity to stress (stress-sensitivity) is associated with, for example, psychotic disorder, depression and borderline personality disorder, and is also present in the first-degree relatives of patients[85]. In addition, research indicates that the familial liability to psychotic disorder is manifested as the momentary subtle expression of 'aberrant salience' in response to stress (psychotic reactivity, for example, subtle perceptual anomalies or paranoid ideas) in the flow of daily life[86]. Furthermore, there is evidence that motivational states, conceptualized as the momentary positive affective responses to daily life positive events, are sensitive to specific molecular genetic variation[87] that predicts similar effects in experimental neuroimaging designs of motivational states[88].

These data indicate that the behavioural expression of liability to psychiatric disorders can be conceptualized as alterations in context-sensitivity expressed as environmental reactivity. Subsequent work has suggested that the origins of environmental reactivity phenotypes, such as stress sensitivity, can be traced to gene–environment interactions occurring over the life course[66]. The implication for aetiological research is that the causes of psychiatric disorders such as psychotic syndrome may be productively studied by focusing on liability phenotypes, such as relatively non-specific alterations in environmental reactivity, rather than on hypothesized specific, confound-free and static disease entities.

Conclusions

The human brain has evolved as a highly context-sensitive system, enabling behavioural flexibility in the face of constantly changing environmental challenges. There is evidence that genetic liability for psychotic syndrome is mediated in part by differential sensitivity to environments of victimization, experience of social exclusion and substances affecting brain functioning, having an impact during development. Given the complexity of the phenotype and evidence of dynamic developmental trajectories, with environmentally sensitive periods, longitudinal research on gene–environment interplay driving variation in behavioural expression of liability, that subsequently may give rise to more severe and more 'co-morbid' expressions of psychopathology and need for care, is required to identify the causes and trajectories of the psychotic syndrome. Examination of differential sensitivity to the environment requires technology to assess directly situated phenotypes indexing dynamic, within-person environmental reactivity as substrate for molecular genetic studies; parallel multidisciplinary translational research, using novel paradigms, may help identify underlying mechanisms and point the way to possible interventions.

1. McGrath, J. et al. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. BMC Med. **2,** 13 (2004).
   Important report, bringing together a massive amount of data, summarizing the epidemiological evidence on the influences of sex, urbanicity and migrant status on the onset of schizophrenia.
2. Perälä, J. et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. Arch. Gen. Psychiatry **64,** 19–28 (2007).
3. van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P. & Krabbendam, L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. Psychol. Med. **39,** 179–195 (2009).
4. van Os, J. & Kapur, S. Schizophrenia. Lancet **374,** 635–645 (2009).
5. Cantor-Graae, E. & Selten, J. P. Schizophrenia and migration: a meta-analysis and review. Am. J. Psychiatry **162,** 12–24 (2005).
6. March, D. et al. Psychosis and place. Epidemiol. Rev. **30,** 84–100 (2008).
   Systematic review on the spatial variation in the distribution of psychotic disorder, indicating an important role for social exposures.
7. Krabbendam, L. & van Os, J. Schizophrenia and urbanicity: a major environmental influence–conditional on genetic risk. Schizophr. Bull. **31,** 795–799 (2005).
8. Kelly, B. D. et al. Schizophrenia and the city: a review of literature and prospective study of psychosis and urbanicity in Ireland. Schizophr. Res. **116,** 75–89 (2010).
9. Keller, M. C., Medland, S. E. & Duncan, L. E. Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. Behav. Genet. **40,** 377–393 (2010).
10. Hutchinson, G. et al. Morbid risk of schizophrenia in first-degree relatives of white and African-Caribbean patients with psychosis. Br. J. Psychiatry **169,** 776–780 (1996).
11. van Os, J., Hanssen, M., Bak, M., Bijl, R. V. & Vollebergh, W. Do urbanicity and familial liability coparticipate in causing psychosis? Am. J. Psychiatry **160,** 477–482 (2003).
12. van Os, J., Rutten, B. P. & Poulton, R. Gene-environment interactions in schizophrenia: review of epidemiological findings and future directions. Schizophr. Bull. **34,** 1066–1082 (2008).
13. Guo, S. W. Gene-environment interaction and the mapping of complex traits: some statistical models and their implications. Hum. Hered. **50,** 286–303 (2000).
14. Polanczyk, G. et al. Etiological and clinical features of childhood psychotic symptoms: results from a birth cohort. Arch. Gen. Psychiatry **67,** 328–338 (2010).
    Prospective study on a very well characterized UK birth cohort, showing alterations in developmental pathways in children with expression of liability for psychotic syndrome in the form of subclinical psychotic experiences.
15. Kendler, K. S. et al. Sources of individual differences in depressive symptoms: analysis of two samples of twins and their families. Am. J. Psychiatry **151,** 1605–1614 (1994).
16. Chen, L. S., Rice, T. K., Thompson, P. A., Barch, D. M. & Csernansky, J. G. Familial aggregation of clinical and neurocognitive features in sibling pairs with and without schizophrenia. Schizophr. Res. **111,** 159–166 (2009).
17. Toulopoulou, T. et al. Substantial genetic overlap between neurocognition and schizophrenia: genetic modeling in twin samples. Arch. Gen. Psychiatry **64,** 1348–1355 (2007).
18. Maric, N. et al. Is our concept of schizophrenia influenced by Berkson's bias? Soc. Psychiatry Psychiatr. Epidemiol. **39,** 600–605 (2004).
19. Dominguez, M. D., Saka, M. C., Lieb, R., Wittchen, H. U. & van Os, J. Early expression of negative/disorganized symptoms predicting psychotic experiences and subsequent clinical psychosis: a 10-year study. Am. J. Psychiatry **167,** 1075–1082 (2010).
20. Dominguez, M. D., Wichers, M., Lieb, R., Wittchen, H. U. & van Os, J. Evidence that onset of clinical psychosis is an outcome of progressively more persistent subclinical psychotic experiences: an 8-year cohort study. Schizophr. Bull. doi:10.1093/schbul/sbp022 (21 May 2009).
21. Read, J., van Os, J., Morrison, A. P. & Ross, C. A. Childhood trauma, psychosis and schizophrenia: a literature review with theoretical and clinical implications. Acta Psychiatr. Scand. **112,** 330–350 (2005).
22. Bendall, S., Jackson, H. J., Hulbert, C. A. & McGorry, P. D. Childhood trauma and psychotic disorders: a systematic, critical review of the evidence. Schizophr. Bull. **34,** 568–579 (2008).
23. Morgan, C. & Fisher, H. Environment and schizophrenia: childhood trauma—a critical review. Schizophr. Bull. **33,** 3–10 (2007).
24. Schreier, A. et al. Prospective study of peer victimization in childhood and psychotic symptoms in a nonclinical population at age 12 years. Arch. Gen. Psychiatry **66,** 527–536 (2009).

25. Arseneault, L. *et al.* Childhood trauma and children's emerging psychotic symptoms: a genetically sensitive longitudinal cohort study. *Am. J. Psychiatry* (in the press).
26. Elklit, A. & Shevlin, M. Female sexual victimization predicts psychosis: a case-control study based on the Danish registry system. *Schizophr. Bull.* doi:10.1093/schbul/sbq048 (20 May 2010).
27. Fisher, H. L. *et al.* Reliability and comparability of psychosis patients' retrospective reports of childhood abuse. *Schizophr. Bull.* doi:10.1093/schbul/sbp103 (23 September 2009).
28. Janssen, I. *et al.* Childhood abuse as a risk factor for psychotic experiences. *Acta Psychiatr. Scand.* **109,** 38–45 (2004).
29. Bourque, F., van der Ven, E. & Malla, A. A meta-analysis of the risk for psychotic disorders among first- and second-generation immigrants. *Psychol. Med.* doi:10.1017/S0033291710001406 (21 July 2010).
    Recent, detailed meta-analytical analysis showing that increased risk of schizophrenia among immigrants persists into the second generation, and that this may be mediated by social context.
30. Bresnahan, M. *et al.* Race and risk of schizophrenia in a US birth cohort: another example of health disparity? *Int. J. Epidemiol.* **36,** 751–758 (2007).
31. Veling, W. *et al.* Ethnic density of neighborhoods and incidence of psychotic disorders among immigrants. *Am. J. Psychiatry* **165,** 66–73 (2008).
32. Boydell, J. *et al.* Incidence of schizophrenia in ethnic minorities in London: ecological study into interactions with environment. *Br. Med. J.* **323,** 1336 (2001).
33. Morgan, C., Charalambides, M., Hutchinson, G. & Murray, R. M. Migration, ethnicity, and psychosis: toward a sociodevelopmental model. *Schizophr. Bull.* doi:10.1093/schbul/sbq051 (30 May 2010).
    Excellent review summarizing current evidence and future perspectives on environmental influences on psychotic disorder, particularly ethnic minority group, integrated in a novel sociodevelopmental hypothesis of psychotic disorder.
34. Selten, J. P. & Cantor-Graae, E. Social defeat: risk factor for schizophrenia? *Br. J. Psychiatry* **187,** 101–102 (2005).
35. Mortensen, P. B., Pedersen, M. G. & Pedersen, C. B. Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? *Psychol. Med.* **40,** 201–210 (2010).
36. Pedersen, C. B. & Mortensen, P. B. Evidence of a dose-response relationship between urbanicity during upbringing and schizophrenia risk. *Arch. Gen. Psychiatry* **58,** 1039–1046 (2001).
37. van Os, J., Driessen, G., Gunther, N. & Delespaul, P. Neighbourhood variation in incidence of schizophrenia. Evidence for person-environment interaction. *Br. J. Psychiatry* **176,** 243–248 (2000).
38. Zammit, S. *et al.* Individuals, schools, and neighborhood: a multilevel longitudinal study of variation in incidence of psychotic disorders. *Arch. Gen. Psychiatry* **67,** 914–922 (2010).
    Fascinating longitudinal study showing how individual-level risk factors may vary from protective to risk-increasing depending on the degree to which they are the norm or the exception in relation to the wider social environment.
39. Morrison, P. D. *et al.* The acute effects of synthetic intravenous Δ^9-tetrahydrocannabinol on psychosis, mood and cognitive functioning. *Psychol. Med.* **39,** 1607–1616 (2009).
40. D'Souza, D. C. *et al.* Delta-9-tetrahydrocannabinol effects in schizophrenia: implications for cognition, psychosis, and addiction. *Biol. Psychiatry* **57,** 594–608 (2005).
    A methodologically rigorous human experimental study on the short-term effects of delta-9-tetrahydrocannabinol (the main psychotropic component of cannabis) on cognitive and clinical phenotypes in schizophrenia.
41. Minozzi, S. *et al.* An overview of systematic reviews on cannabis and psychosis: discussing apparently conflicting results. *Drug Alcohol Rev.* **29,** 304–317 (2010).
42. Ferdinand, R. F. *et al.* Cannabis use predicts future psychotic symptoms, and vice versa. *Addiction* **100,** 612–618 (2005).
43. Fergusson, D. M., Horwood, L. J. & Swain-Campbell, N. R. Cannabis dependence and psychotic symptoms in young people. *Psychol. Med.* **33,** 15–21 (2003).
44. Veling, W., Mackenbach, J. P., van Os, J. & Hoek, H. W. Cannabis use and genetic predisposition for schizophrenia: a case-control study. *Psychol. Med.* **38,** 1251–1256 (2008).
45. Genetic Risk and Outcome in Psychosis (GROUP) Investigators. Evidence that familial liability for psychosis is expressed as differential sensitivity to cannabis: an analysis of patient-sibling and sibling-control pairs. *Arch. Gen. Psychiatry* doi:10.1001/archgenpsychiatry.2010.132 (4 October 2010).
46. van Winkel, R., Genetic Risk and Outcome in Psychosis (GROUP) Investigators. Family-based analysis of genetic variation underlying psychosis-inducing effects of cannabis: sibling analysis and proband follow-up. *Arch. Gen. Psychiatry* (in the press).
    Elegant gene–environment interaction study, combining a hypothesis-based with a more systematic approach, suggestive of interaction of cannabis with specific molecular markers of genetic variation.
47. Di Forti, M. *et al.* High-potency cannabis and the risk of psychosis. *Br. J. Psychiatry* **195,** 488–491 (2009).
48. Cannon, M., Jones, P. B. & Murray, R. M. Obstetric complications and schizophrenia: historical and meta-analytic review. *Am. J. Psychiatry* **159,** 1080–1092 (2002).
49. van Os, J., Pedersen, C. B. & Mortensen, P. B. Confirmation of synergy between urbanicity and familial liability in the causation of psychosis. *Am. J. Psychiatry* **161,** 2312–2314 (2004).
50. Frith, C. D. & Corcoran, R. Exploring 'theory of mind' in people with schizophrenia. *Psychol. Med.* **26,** 521–530 (1996).

51. Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D. & Wolpert, D. M. Evidence for sensory prediction deficits in schizophrenia. *Am. J. Psychiatry* **162,** 2384–2386 (2005).
52. Mason, O. J. & Brady, F. The psychotomimetic effects of short-term sensory deprivation. *J. Nerv. Ment. Dis.* **197,** 783–785 (2009).
53. Galdos, M. *et al.* Affectively salient meaning in random noise: a task sensitive to psychosis liability. *Schizophr. Bull.* doi:10.1093/schbul/sbq029 (1 April 2010).
54. Peterson, C. C. & Siegal, M. Deafness, conversation and theory of mind. *J. Child Psychol. Psychiatry* **36,** 459–474 (1995).
55. Colvert, E. *et al.* Do theory of mind and executive function deficits underlie the adverse outcomes associated with profound early deprivation?: findings from the English and Romanian adoptees study. *J. Abnorm. Child Psychol.* **36,** 1057–1068 (2008).
56. David, A., Malmberg, A., Lewis, G., Brandt, L. & Allebeck, P. Are there neurological and sensory risk factors for schizophrenia? *Schizophr. Res.* **14,** 247–251 (1995).
57. Stefanis, N., Thewissen, V., Bakoula, C., van Os, J. & Myin-Germeys, I. Hearing impairment and psychosis: a replication in a cohort of young adults. *Schizophr. Res.* **85,** 266–272 (2006).
58. Malaspina, D. *et al.* Traumatic brain injury and schizophrenia in members of schizophrenia and bipolar disorder pedigrees. *Am. J. Psychiatry* **158,** 440–446 (2001).
59. Barkus, E. & Murray, R. M. Substance use in adolescence and psychosis: clarifying the relationship. *Annu. Rev. Clin. Psychol.* **6,** 365–389 (2010).
60. Martín-Rodríguez, J. F. & León-Carrión, J. Theory of mind deficits in patients with acquired brain injury: a quantitative review. *Neuropsychologia* **48,** 1181–1191 (2010).
61. Homer, B. D. *et al.* Methamphetamine abuse and impairment of social functioning: a review of the underlying neurophysiological causes and behavioral implications. *Psychol. Bull.* **134,** 301–310 (2008).
62. Fett, A. K. *et al.* The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neurosci. Biobehav. Rev.* doi:10.1016/j.neubiorev.2010.07.001 (8 July 2010).
63. Arnsten, A. F. Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Rev. Neurosci.* **10,** 410–422 (2009).
64. Kaschube, M., Wolf, F., Geisel, T. & Lowel, S. Genetic influence on quantitative features of neocortical architecture. *J. Neurosci.* **22,** 7206–7217 (2002).
65. Glaser, J. P., van Os, J., Portegijs, P. J. & Myin-Germeys, I. Childhood trauma and emotional reactivity to daily life stress in adult frequent attenders of general practitioners. *J. Psychosom. Res.* **61,** 229–236 (2006).
66. Wichers, M. *et al.* Mechanisms of gene-environment interactions in depression: evidence that genes potentiate multiple sources of adversity. *Psychol. Med.* **39,** 1077–1086 (2009).
67. Lieberman, J. A., Sheitman, B. B. & Kinon, B. J. Neurochemical sensitization in the pathophysiology of schizophrenia: deficits and dysfunction in neuronal regulation and plasticity. *Neuropsychopharmacology* **17,** 205–229 (1997).
68. Carboni, E. *et al.* Prenatal restraint stress: an *in vivo* microdialysis study on catecholamine release in the rat prefrontal cortex. *Neuroscience* **168,** 156–166 (2010).
69. Howes, O. D. & Kapur, S. The dopamine hypothesis of schizophrenia: version III–the final common pathway. *Schizophr. Bull.* **35,** 549–562 (2009).
70. Thomas, D. Gene-environment-wide association studies: emerging approaches. *Nature Rev. Genet.* **11,** 259–272 (2010).
71. Darroch, J. Biologic synergism and parallelism. *Am. J. Epidemiol.* **145,** 661–668 (1997).
72. Caspi, A. & Moffitt, T. E. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nature Rev. Neurosci.* **7,** 583–590 (2006).
73. Schmidt-Kastner, R., van Os, J., Steinbusch, H. W. M. & Schmitz, C. Gene regulation by hypoxia and the neurodevelopmental origin of schizophrenia. *Schizophr. Res.* **84,** 253–271 (2006).
74. Rutten, B. P. & Mill, J. Epigenetic mediation of environmental influences in major psychotic disorders. *Schizophr. Bull.* **35,** 1045–1056 (2009).
75. Keen, D. V., Reid, F. D. & Arnone, D. Autism, ethnicity and maternal immigration. *Br. J. Psychiatry* **196,** 274–281 (2010).
76. Gardener, H., Spiegelman, D. & Buka, S. L. Prenatal risk factors for autism: comprehensive meta-analysis. *Br. J. Psychiatry* **195,** 7–14 (2009).
77. Miller, B. *et al.* Meta-analysis of paternal age and schizophrenia risk in male versus female offspring. *Schizophr, Bull.* doi:10.1093/schbul/sbq011 (25 February 2010).
78. McMillan, K. A., Enns, M. W., Cox, B. J. & Sareen, J. Comorbidity of axis I and II mental disorders with schizophrenia and psychotic disorders: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Can. J. Psychiatry* **54,** 477–486 (2009).
79. Hanssen, M. *et al.* How psychotic are individuals with non-psychotic disorders? *Soc. Psychiatry Psychiatr. Epidemiol.* **38,** 149–154 (2003).
80. Weiser, M. *et al.* Cognitive performance of male adolescents is lower than controls across psychiatric disorders: a population-based study. *Acta Psychiatr. Scand.* **110,** 471–475 (2004).
81. Weiser, M. *et al.* Subtle cognitive dysfunction in nonaffected siblings of individuals affected by nonpsychotic disorders. *Biol. Psychiatry* **63,** 602–608 (2008).
    Innovative study reporting familial clustering of cognitive dysfunction in psychiatric disorders that cuts across diagnostic psychiatric entities.
82. Argyropoulos, S. V. *et al.* Twins discordant for schizophrenia: psychopathology of the non-schizophrenic co-twins. *Acta Psychiatr. Scand.* **118,** 214–219 (2008).
83. Goodwin, R. D., Fergusson, D. M. & Horwood, L. J. Neuroticism in adolescence and psychotic symptoms in adulthood. *Psychol. Med.* **33,** 1089–1097 (2003).
84. Rodgers, B. Behaviour and personality in childhood as predictors of adult psychiatric disorder. *J. Child Psychol. Psychiatry* **31,** 393–414 (1990).

85. Myin-Germeys, I. *et al.* Experience sampling research in psychopathology: opening the black box of daily life. *Psychol. Med.* **39,** 1533–1547 (2009).
86. Myin-Germeys, I., Delespaul, P. & van Os, J. Behavioural sensitization to daily life stress in psychosis. *Psychol. Med.* **35,** 733–741 (2005).
87. Wichers, M. *et al.* The catechol-*O*-methyl transferase Val[158]Met polymorphism and experience of reward in the flow of daily life. *Neuropsychopharmacology* **33,** 3030–3036 (2008).
88. Dreher, J. C., Kohn, P., Kolachana, B., Weinberger, D. R. & Berman, K. F. Variation in dopamine genes influences responsivity of the human reward system. *Proc. Natl Acad. Sci. USA* **106,** 617–622 (2009).
89. Weiser, M. *et al.* Social and cognitive functioning, urbanicity and risk for schizophrenia. *Br. J. Psychiatry* **191,** 320–324 (2007).
90. Habets, P. *et al.* Reduced cortical thickness as an outcome of differential sensitivity to environmental risks in schizophrenia. *Biol. Psychiatry* doi:10.1016/j.biopsych.2010.08.010 (16 October 2010).
91. Grace, A. A., Floresco, S. B., Goto, Y. & Lodge, D. J. Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci.* **30,** 220–227 (2007).
    Comprehensive review summarizing the regulation of mesolimbic dopamine signalling in the integration of top-down cognitive control and bottom-up input.
92. Mueller, S. C. *et al.* Early-life stress is associated with impairment in cognitive control in adolescence: an fMRI study. *Neuropsychologia* **48,** 3037–3044 (2010).
93. Pruessner, J. C., Champagne, F., Meaney, M. J. & Dagher, A. Dopamine release in response to a psychological stress in humans and its relationship to early life maternal care: a positron emission tomography study using [11C]raclopride. *J. Neurosci.* **24,** 2825–2831 (2004).
    PET imaging study showing long-term influence of parental care during early life on dopamine release in the ventral striatum under psychosocial stress in adulthood.
94. Sheu, Y. S., Polcari, A., Anderson, C. M. & Teicher, M. H. Harsh corporal punishment is associated with increased T2 relaxation time in dopamine-rich regions. *Neuroimage* **53,** 412–419 (2010).
95. Taylor, S. E., Eisenberger, N. I., Saxbe, D., Lehman, B. J. & Lieberman, M. D. Neural responses to emotional stimuli are associated with childhood family stress. *Biol. Psychiatry* **60,** 296–301 (2006).
96. Jager, G., Block, R. I., Luijten, M. & Ramsey, N. F. Cannabis use and memory brain function in adolescent boys: a cross-sectional multicenter functional magnetic resonance imaging study. *J. Am. Acad. Child Adolesc. Psychiatry* **49,** 561–572 (2010).
97. Nestor, L., Hester, R. & Garavan, H. Increased ventral striatal BOLD activity during non-drug reward anticipation in cannabis users. *Neuroimage* **49,** 1133–1143 (2010).
98. Nestor, L., Roberts, G., Garavan, H. & Hester, R. Deficits in learning and memory: parahippocampal hyperactivity and frontocortical hypoactivity in cannabis users. *Neuroimage* **40,** 1328–1339 (2008).
99. Yücel, M. *et al.* Regional brain abnormalities associated with long-term heavy cannabis use. *Arch. Gen. Psychiatry* **65,** 694–701 (2008).

natureINSIGHT

GLIA

Cover illustration
Mouse astrocyte (white), blood vessels and neuronal dendrites imaged with a Gatan 3View (E. Bushong/ M. Terada/M. Ellisman).

Ever since their discovery more than 150 years ago, glial cells have been defined by what they couldn't do. Above all, they have lacked the ability to communicate with other cells through fast electrical and chemical signals — action potentials running down axons and transmitter release at synapses. This has been the exclusive property of neurons, their arch rivals in the battle for researchers' attention — and funding. And because action potentials remain the dominant currency of information processing in the brain, most neuroscientists still see glia as second-class supporting material — the brain's 'glue'.

But sure enough, each year brings further exceptions to such a strict division of labour. In 2008, for example, a subclass of glial cell was reported to fire bona fide action potentials. And a sizeable fraction of the glial research community entertains the radical view that evoked release of neurotransmitters by glial cells — 'gliotransmission' — must contribute to the information-processing power of the brain (see News Feature, page 160).

Neuron-envy should now become a thing of the past. Recent years have seen an explosion of new findings demonstrating that glial cells play an irreplaceable part in all aspects of brain function. And with this collection of reviews, we want to illustrate some of the most vibrant sectors of glial research, and the major impact that glia have on issues as diverse as brain development, neuronal plasticity and the control of cerebral blood flow, which makes functional brain imaging possible.

As every article in this supplement also amply demonstrates, glia have a central role in many diseases of the nervous system. And the numerous new molecular markers and tracing techniques discussed here offer just as many fresh opportunities for research and therapeutic intervention.

We hope this supplement will make it clear that a brain cell doesn't need to be a neuron to enjoy a lot of the action and to offer great potential. At the very least, glial cells should now be defined by what they can do.
Tanguy Chouard and Noah Gray
Senior Editors

CONTENTS

REVIEW

Developmental genetics of vertebrate glial–cell specification

David H. Rowitch[1,2,3] & Arnold R. Kriegstein[4]

Oligodendrocytes and astrocytes are macroglial cells of the vertebrate central nervous system. These cells have diverse roles in the maintenance of neurological function. In the embryo, the genetic mechanisms that underlie the specification of macroglial precursors *in vivo* appear strikingly similar to those that regulate the development of the diverse neuron types. The switch from producing neuronal to glial subtype-specific precursors can be modelled as an interplay between region-restricted components and temporal regulators that determine neurogenic or gliogenic phases of development, contributing to glial diversity. Gaining insight into the developmental genetics of macroglia has great potential to improve our understanding of a variety of neurological disorders in humans.

Glia make up 10–20% of the cells in the *Drosophila* nervous system and at least 50% of the cells in the human brain. These findings indicate that glial-cell function is crucial for the increase in complexity of neurological function that has emerged during evolution. The principal types of macroglial cell — astrocytes and oligodendrocytes — are derived from the neuroepithelium and are found throughout the mature central nervous system (CNS). By contrast, the other group of glia — microglia — are mesodermal (more specifically, haematopoietic) in origin.

Astrocytes provide structural support, regulate water balance and ion distribution, and maintain the blood–brain barrier. They also participate in cell–cell signalling by regulating calcium flux, releasing D-serine, producing neuropeptides and modulating synaptic transmission. In the CNS, myelin provides insulation for neuronal axons and allows saltatory conduction through the formation of nodes of Ranvier. Some invertebrates have ensheathing glia that produce components of myelin[1] and even cytoplasmic extrusions and node-of-Ranvier-like structures[2], but oligodendrocytes capable of forming compact myelin are present in all jawed vertebrates. Moreover, oligodendrocyte precursor cells (OPCs) in the mammalian brain form synapses with neurons, suggesting an even greater degree of complexity in the interactions between neurons and oligodendroglia[3].

The diverse macroglial-cell types can be distinguished both by their morphological characteristics and by their expression of various markers (Fig. 1). Mature oligodendrocytes of the white matter express a variety of myelin markers, including myelin basic protein, proteolipid protein 1 (PLP1) and adenomatous polyposis coli (APC) protein. OPCs, by contrast, express platelet-derived growth-factor receptor-α, the transcription factor SOX10 and the proteoglycan NG2 (also known as CSPG4). These cells maintain proliferative and migratory competence during development and in the adult, and are early responders to injury (Fig. 1). Mature astrocytes can be divided into two categories: fibrous astrocytes and protoplasmic astrocytes. Fibrous astrocytes populate the white matter and typically have a 'star-like' appearance with dense glial filaments that can be stained with the intermediate filament marker glial fibrillary acidic protein (GFAP). Protoplasmic astrocytes are found in the grey matter, have more irregular, 'bushy', processes than do fibrous astrocytes and typically have few glial filaments. These cells come into contact with and ensheath synapses by extending thousands of thin processes, some of which also contact blood vessels.

This review describes recent advances in understanding the developmental genetics that underlies macroglial cell-type specification in vertebrates. In the past decade, it has become clear that the way in which glia derived from the neuroepithelium are specified follows similar rules to the specification of the various neuron types. There is now enough evidence to make the case that glial-cell specification in the embryo is regulated according to a 'segmental template', with the developmental programs for oligodendrocytes and astrocytes being partly independent.

Overview of stem-cell precursors for macroglia

All neurons and macroglia in the developing CNS are derived from neuroepithelial cells that line the cerebral ventricles and spinal canal. At about embryonic day (E) 9–10 in mice, progressive waves of neurogenesis begin at caudal regions of the spinal cord and proceed rostrally, as well as along ventral–dorsal and lateral–medial gradients in the brain. Radial glial cells are the primary progenitor cells at embryonic stages of neurogenesis[4–6], and like the neuroepithelial cells from which they are derived, they line the forebrain ventricles and spinal canal, maintain apical–basal polarity, and undergo interkinetic nuclear migration in association with cell-cycle progression.

Radial glia differentiate into neurons and macroglia, as has been shown by tracking cells using Cre–*loxP* fate mapping with Cre expressed under the control of the brain lipid-binding protein (*Blbp*) gene promoter[7]. Time-lapse imaging of radial glia shows that they often undergo asymmetrical self-renewal division to produce neurons or intermediate progenitor cells[8–11]. Oligodendrocyte precursors and ependymal cells also derive from radial glial cells, but whether intermediate progenitors are involved is uncertain. Intermediate progenitors are the main proliferative cells of the subventricular zones of the embryonic telencephalon[8,10] and seem to be restricted to producing neurons or glia (Fig. 1).

The *in vivo* potential of neuroepithelial cells and radial glia becomes regionally restricted through the action of organizing signals such as sonic hedgehog (SHH), fibroblast growth factors (FGFs), WNTs and bone morphogenetic proteins (BMPs), all of which provide positional information through morphogen gradients in the dorsal–ventral, anterior–posterior and medial–lateral axes (discussed below in 'Embryonic pattern formation and gliogenesis') .

[1]Howard Hughes Medical Institute, [2]Department of Pediatrics, [3]Department of Neurological Surgery, [4]Department of Neurology, University of California, San Francisco, and the Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, 513 Parnassus Avenue, San Francisco, California 94143-0525, USA.
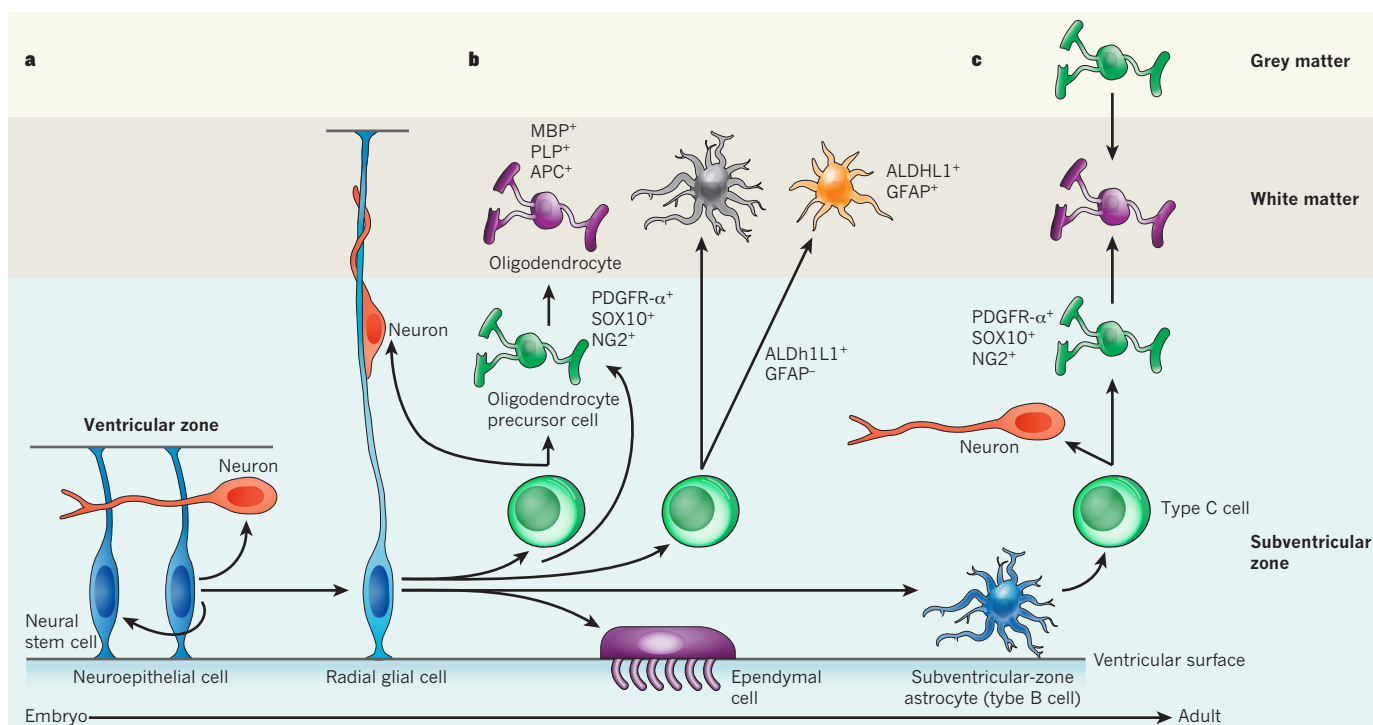
Figure 1 | Patterns of gliogenesis in embryonic and adult progenitor zones. The progression from the embryo to the adult is shown from left to right (**a** to **c**). Black arrows indicate self-renewal or differentiation from one cell type to another. Markers of macroglia and their precursors are listed. **a**, Self-renewing neuroepithelial cells line the ventricles throughout the neuraxis at the stages of neural tube closure. These cells may generate some neurons. Neuroepithelial cells are transformed into radial glial cells as neurogenesis begins. **b**, Radial glia produce intermediate progenitor cells and oligodendrocyte precursor cells (OPCs), which in turn produce neurons and oligodendrocytes, respectively. Radial glia can also become astrocytes, as well as producing intermediate progenitors that expand in number before producing astrocytes. Protoplasmic astrocytes and fibrous astrocytes might arise from common or independent progenitors. Radial glia also produce ependymal cells. **c**, In adults, oligodendrocytes are produced by two independent pathways: type B cells in the cortical subventricular zone produce transit-amplifying cells (known as type C cells), which in turn produce OPCs as well as neurons. The OPCs subsequently generate oligodendrocytes, and OPCs that are already resident in the grey matter also produce oligodendrocytes. ALDH1L1, aldehyde dehydrogenase 1 family, member L1; APC, adenomatous polyposis coli; GFAP, glial fibrillary acidic protein; MBP, myelin basic protein; PDGFR-α, platelet-derived growth-factor receptor-α; PLP, proteolipid protein 1. All green cells are intermediate progenitors, with type C cells being a subset of these, and all blue cells are neural stem cells (even though each blue cell is a different type).

Embryonic pattern formation and gliogenesis

A fundamental question in developmental neurobiology is how a relatively simple and undifferentiated neuroepithelium in the embryo can give rise to the remarkable cellular diversity and specialization of the mature CNS. It is now clear that both spatial and temporal mechanisms operate to generate diverse neuron and glial-cell types. Patterning along the neuraxis leads to segmentation of the neuroepithelium into progenitor domains (denoted p0, p1 p2, p3 and pMN) for distinct neuron types. On the basis of recent studies, this model can be extended to macroglia, but important questions remain about the role of the ultimate environment to further shape key characteristics of macroglia. In this section, we summarize how domains for neuronal and glial progenitors are established in the embryo.

Embryonic spinal cord

Organizing signals are required for pattern formation in the embryonic neural tube. SHH is a secreted protein that is essential for the activity of organizing structures of the ventral midline, such as the notochord and floor plate. Full-length SHH undergoes autoproteolysis and lipid modifications, which are crucial for its long-range signalling properties in the neural tube. Ventral SHH-mediated signalling is antagonized by patched, GLI3, HIP1 and dorsal BMP- and WNT-mediated signalling[12,13]. Together, these molecules establish a gradient of morphogenic activity that confers positional identity in the dorsal–ventral axis, resulting in generation of neural diversity.

An important function of SHH-mediated signalling is to regulate the expression of transcription factors that demarcate unique progenitor regions in the ventral neural tube[12] (Fig. 2a). The subsequent sharpening and maintenance of domain boundaries depends on cross-regulatory interactions: for example, repression of expression of the homeobox protein IRX3 by oligodendrocyte transcription factor 2 (OLIG2), and NKX2.2–PAX6 reciprocal repression (Fig. 2b). Combinatorial interactions within a given progenitor domain regulate its identity and, consequently, neuron type specification and further diversification[12].

SHH-mediated signalling is both necessary and sufficient for oligodendrocyte production in the spinal cord, another parallel with the development of motor neurons[14]. Both the concentration and timing of SHH exposure is important for establishing cell fate[15]. Indeed, Orentas et al.[14] found that SHH activity is required up until the time of OPC specification, at stage 24 in the chick (about E12.5 in mice). By contrast, later stages of OPC maturation are SHH independent[14,16], which is consistent with the normal sequence of OPC migration away from the source of SHH in the ventral midline (Fig. 3).

Functional analyses of transcription-factor-encoding genes have provided insight into the distinct gliogenic domains for oligodendrocytes and astrocytes in the ventral neural tube. *Olig1/2* null mutant mice show a failure in the development of motor neurons and OPCs in the spinal cord, as well as of all oligodendrocytes in the brain[17,18]. Data indicating that OLIG proteins function as transcriptional repressors show that the transcriptional targets of OLIG proteins might themselves be antagonists of motor neuron or oligodendrocyte development, such that OLIG proteins promote a neuron or glial-cell fate by repressing the repressors of that fate[19]. This mechanistic parallel between motor neuron and OPC

a Patterning

b Neurogenesis and gliogenesis

c

d
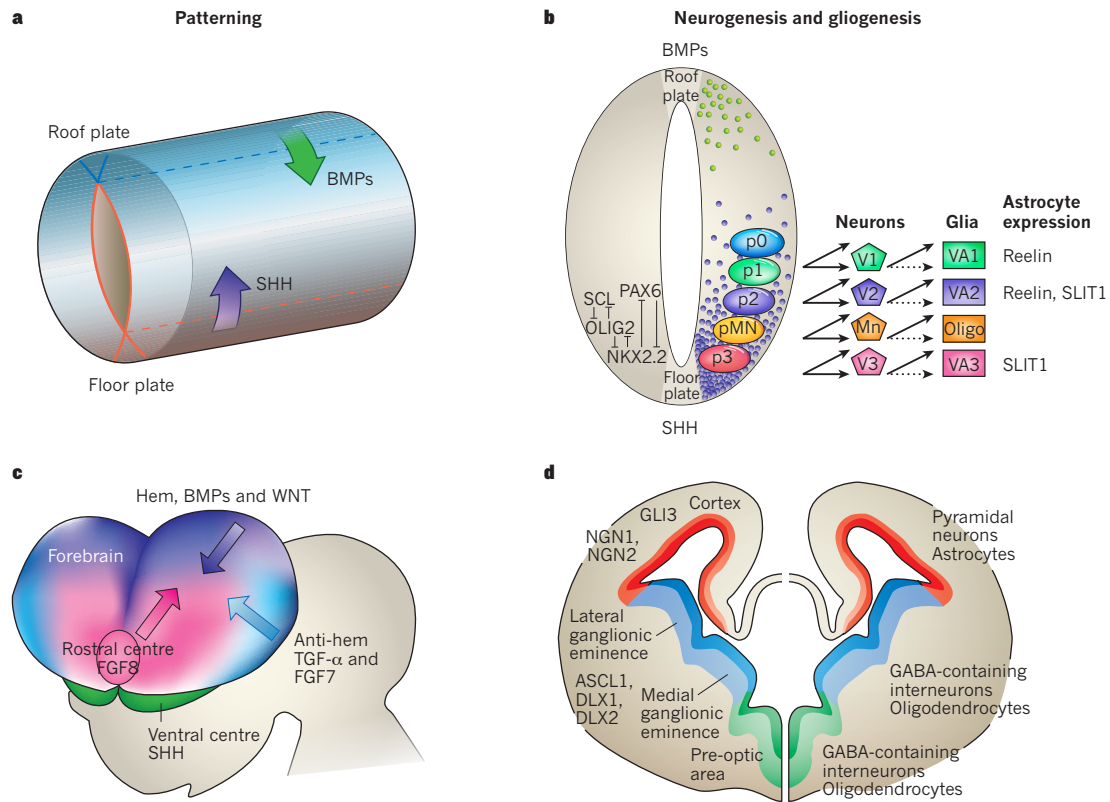
Figure 2 | Patterning of the neural tube generates unique domains for neuronal and glial progenitors. a, The primitive neuroepithelium of the neural tube is patterned by organizing signals. These signals emanate from the ventral floor plate (such as SHH, purple) and roof plate (BMPs and WNTs, green). **b**, A cross-sectional view of the neural tube is shown. Progenitors of motor neurons and interneurons are formed within distinct regionally restricted domains of the ventral neural tube: the p0, p1, p2 and p3 domains for interneuron subtypes, and the pMN domain for motor neurons. Dorsal domains are also similarly parcelled (not shown). Signalling mediated by SHH (gradient denoted by purple circles) regulates the expression of transcription factors (for example, NKX2.2, OLIG2, PAX6 and SCL) in the ventral neural tube. The interactions of these factors sharpen and maintain the domain boundaries. Embryonic OPCs are derived mainly from the pMN domain. OPCs are recognized by expression of PDGFR-α, SOX10 and NG2. Three astrocyte subtypes have been identified: VA1 astrocytes (which express PAX6 and reelin, derived from p1) are the most dorsal; VA3 astrocytes (which express NKX6.1 and SLIT1, derived from p3) are the most ventral; and VA2

astrocytes (which express PAX6, NKX6.1, reelin and SLIT1, derived from p2) are located in an intermediate white-matter domain. **c**, Organizing centres of the forebrain are shown. These include the cortical hem (purple), which is a dorsal source of BMPs and WNTs; a ventral centre (green), which is a source of SHH; and rostral (pink) and anti-hem (blue) regions, which are sources of growth factors such as FGF8, and transforming growth factor-α (TGF-α) and FGF7, respectively. **d**, A coronal view of the embryonic (~E14.5) forebrain showing its division into dorsal and ventral regions that are specialized for producing different neuron and glial-cell types. The dorsal region includes the cortex, a source of pyramidal neurons and astrocytes. The ventral region includes the lateral and medial ganglionic eminences and the pre-optic area, which are sources of GABA (γ-aminobutyric acid)-containing interneurons and oligodendrocytes. The green, blue and red shaded areas represent the pre-optic area, medial/lateral ganglionic eminences and neocortex, respectively (see Fig. 3). Transcription factors that are associated with dorsal (NGN1, NGN2, GLI3) and ventral (ASCL1, DLX1, DLX2) patterning and cell fate specification are indicated. NGN, neurogenin.

development, however, does not indicate that these cells originate from a common bipotent progenitor. Indeed, several lines of evidence indicate that the pMN domain (where motor neurons form) comprises independently segregating neuroblasts and glioblasts that are sequentially specified (Box 1).

Astrocyte development was increased in *Olig1/2* mutants, as indicated by more cells expressing GFAP. Such *Olig1/2* double mutants and *Olig2* single mutants lack the pMN domain and, instead, develop a ventrally expanded p2 domain and have more interneurons of the V2 class and more astrocytes. In other words, astrocytes develop despite a genetic deletion that eliminates the ultimate precursor population for OPCs, a finding that is incompatible with the bipotent glial-cell-restricted precursor model of development (Box 1).

Given these findings, ventral neural-tube astrocytes and oligodendrocytes seem to develop along mutually exclusive paths, but what would be the underlying molecular mechanisms of this? One possibility is that OLIG2 represses astrocyte fate by promoting an alternative fate. Alternatively, OLIG2 might repress a 'pro-astrocytic' transcriptional program in an adjacent progenitor domain. In support of the latter model, a basic

helix–loop–helix (bHLH) code operates in the neural tube, and this is required for astrocyte differentiation within the p2 domain[20]. Whereas deletion of *Olig2* expands the p2 domain ventrally, generating astrocytes instead of oligodendrocytes[18], deletion of p2-associated *Scl* (also known as *Tal1*) resulted in dorsalized OLIG2 expression, increasing oligodendrocyte production at the expense of astrocytes. Data showing that cross-repressive interactions between the bHLH transcription factors SCL and OLIG2 are required to maintain astrocyte generation suggested a model whereby astrogenesis might generally take place in restricted regions of the neural tube.

This idea of a 'segmental model' for glial-cell type specification is further supported by work by Hochstim *et al.*[21], who showed that combinatorial expression of PAX6 and NKX6.1 specified three molecularly distinct subtypes of ventral astrocytes (termed VA1, VA2 and VA3), which were identified on the basis of their expression of *Slit1* and *Reelin*, which encode axon and neuronal migration factors. These astrocyte subtypes show dorsal–ventral positional identity in white matter, which mirrors the arrangement of progenitors in the p1, p2 and p3 domains. NKX2.2 seems to be required for SLIT1 expression in ventral astrocyte precursors,

BOX 1

Investigating the potential of glial precursors

Here we consider proposed models for glial-cell lineage progression, caveats to these models and new directions that might better explain the nature of glial precursor cells *in vivo*.

Are there bipotent precursors for glial-cell types *in vivo*?

The developmental potential of the neural progenitor cells that give rise to astrocytes and oligodendrocytes remains the subject of extensive debate[84]. Seminal studies by Raff and colleagues identified two types of astrocyte precursor *in vitro* (type 1 and type 2 astrocytes*)* and indicated that type 2 astrocytes and oligodendrocytes developed from a common 'O-2A' precursor[85]. Subsequently, others reported a glial-restricted precursor (GRP) cell, which was found to be competent to generate all three glial-cell types on the basis of *in vitro* analysis of cultured cells and their potential after transplantation[86,87].

In vivo studies, however, do not strongly support the model of normal development through such a bipotent oligodendrocyte–astrocyte (O-A) GRP cell. Subjecting glial precursors *in vivo* (reviewed in ref. 88) to retroviral fate mapping (an approach that targets proliferating cells) indicates that mixed astrocyte and oligodendrocyte clones were rarely or never observed in embryos. This finding is in keeping with data from *Olig2*-knockout animals and from animals in which the pMN domain has been ablated, both of which completely lack embryonic OPCs but have unaffected astrocyte production[89]. This result indicates that OPCs and astrocytes develop in different regions of the neural tube by independent mechanisms (see Fig. 2b). By contrast, the Goldman[90] and Parnavelas[91] groups have shown that bipotent O-A clones could be labelled with retrovirus in the neonatal forebrain. However, such cells represent only 10–15% of glia-generating precursors. Together, these data indicate that the bipotent O-A GRP is only a minor pathway in gliogenesis, at a transient neonatal stage of brain development. A common bipotent O-A precursor might have a larger role in the setting of injury[92,93].

Another alternative model has also been put forward. The lack of motor neurons and OPCs in *Olig2*-null animals might be explained by the existence of a restricted bipotent motor-neuron–OPC precursor cell. However, several findings indicate, instead, that the pMN domain comprises independently segregating neuroblasts and glioblasts rather than a bipotent precursor[27]. Collectively, the studies above argue against glial-cell development through restricted bipotent O-A or neuron–oligodendrocyte precursors.

Restriction of neuron versus glial precursor potential

In vertebrate systems, many lines of evidence are consistent with a model of general restriction of CNS precursor cells to the production of neurons at early stages, followed by a later phase of glial-cell production (see the section 'The switch from neuron to glial-cell production'). In this way, a precursor at the phase of glial-cell production might be called 'glial-restricted' in the sense that it cannot normally produce neurons. The actions of signalling pathways (for example, Delta-like–NOTCH signalling) and the 'pro-glial-cell' transcription factors (for example, SOX9 and nuclear factor I (NFI) proteins) are generally required in all dorsal–ventral domains of the neural tube for the transition to glial-cell production. Such general mechanisms must work

together with region-specific factors to establish different glial-cell types (oligodendrocytes or astrocytes) in particular domains (Fig. 2). Finally, in addition to such 'active' programs of glial-cell type specification, the default mechanism by which radial glia become astrocytes (Fig. 1) is predicted to contribute astrocytes to all domains.

Phases of neuron and glial-cell production from progenitors might alternate. In invertebrates, the production of neuroblasts and glioblasts from a common progenitor can switch back and forth[34]. Type B cells of the adult rodent subventricular zone can give rise to neurons and/or oligodendrocytes[94] at different times in response to environmental cues[95]; however, further work is needed to establish whether this takes place through a common bipotent precursor or reflects segregation of independent neuron- and glial-cell-dedicated type B cells.

In summary, published studies indicate that the classic bipotent glial-restricted precursor defined in *in vitro* studies is unlikely to be the ancestor of most glia *in vivo*. A challenge for developmental genetics is to identify the precise mechanisms that account for astrocyte type specification and lineage progression.

New genetic tools for investigating the astrocyte lineage

Newer fate-mapping approaches might help to address further the issues of glial-cell lineage progression. These approaches include constitutive and inducible recombinase technologies (Flp and Φ31)[96], in combination with specialized reporter systems such as mosaic analysis with double markers (MADM)[97] and Brainbow[98]. These technologies allow for more precise targeting of progenitor populations, for example using two types of recombinase under the control of different yet overlapping enhancers, or using reporters that can be activated in a very few cells, allowing clonal origins to be assessed *in vivo*. However, a major technical hurdle is identifying regulatory sequences that will drive expression of the recombinase specifically in astrocytes and their intermediate progenitor cells. Although human *GFAP*-regulatory sequences have been used extensively in transgenic mice, GFAP is expressed by a heterogeneous group of cells in the CNS. Therefore, careful attention must be paid to the precise expression characteristics of available transgenic mice that use the *GFAP* enhancer in such systems. Other markers of astrocytes and/or their precursors have been reported, including FGF receptor 3 (ref. 99), GLAST (also known as SLC1A3)[100], fatty-acid-binding protein 7 (ref. 101), brain lipid-binding protein (BLBP)[7], SOX9 (ref. 102) and NFIA and NFIB[36]. However, some of these markers (for example, BLBP) are also expressed during neurogenic stages and therefore do not exclusively mark ventricular-zone cells committed to the astrocyte lineage. Other markers are also expressed in non-astrocyte lineages (for example, FGFR3 in oligodendrocytes). Recently, gene-expression profiling of purified astrocyte populations[103,104] has uncovered several potential new markers, such as the gene encoding the folate metabolic enzyme ALDH1L1 (refs 103, 105). Clearly, a priority for the field is identification of markers that are specific for key stages of astrocyte development throughout the CNS, as well as possibly heterogeneous astrocyte precursors and mature astrocytes.

independent of its function to repress PAX6 from the p3 domain[22].

In summary, it is clear that the diverse astrocytes and oligodendrocytes of the ventral neural tube are generated according to the domain organization of the neural. It remains to be seen whether the functional heterogeneity of astrocytes is similarly regulated and whether this 'segmental model' applies to gliogenesis in the dorsal regions of the spinal cord and in the rostral CNS.

Neuroepithelium of the forebrain

As indicated in Fig. 2d, forebrain excitatory pyramidal cells arise in the dorsal telencephalon, whereas inhibitory interneurons arise in the ventral telencephalon. Moreover, different classes of interneuron are derived from specific progenitor domains within the ventral telencephalon. How is such diversity achieved? Similar to the spinal cord, patterning centres located at key points along the neuraxis induce the

a Forebrain

b Neural tube and spinal cord
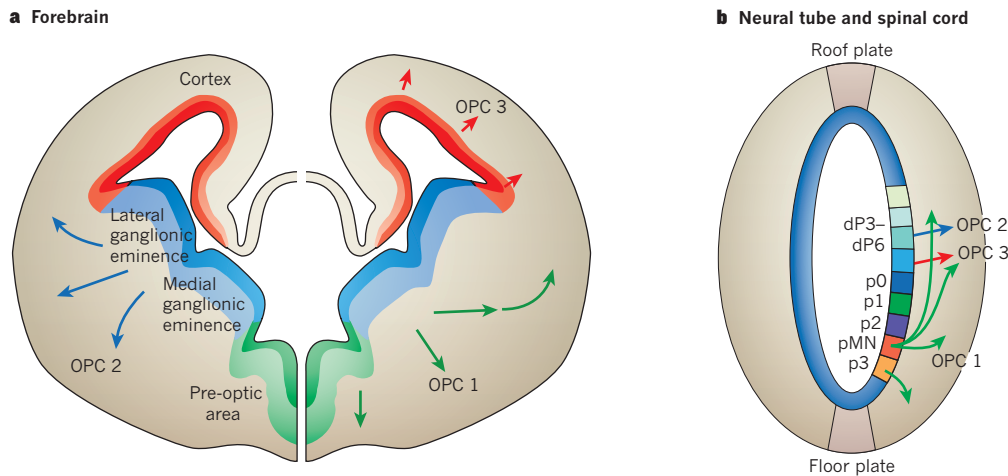
Figure 3 | Multiple waves of oligodendrocyte production in the mammalian CNS. **a**, Three sequential waves of OPCs (OPC 1, OPC 2 and OPC 3) are generated from different regions of the forebrain ventricular zone: OPC 1 (green arrows) arises from *Nkx2.1*-expressing precursors in the medial ganglionic eminence, starting at E12.5; OPC 2 (blue arrows) arises from precursors expressing the homeobox gene *Gsx2* in the lateral and medial ganglionic eminences, starting at E15.5; and OPC 3 (red arrows) arises from precursors expressing the homeobox gene *Emx1* in the cortex, starting at birth. Work from the Richardson lab[70] shows that the OPC 1 population is replaced by later waves[70]. **b**, Two distinct waves of OPCs emanate from the ventral region and the dorsal region of the spinal cord in the embryo and fetus, respectively. A third wave begins after birth. In the first wave (OPC 1, green arrows), ventral OPCs arise from *Olig2*-

expressing progenitor cells in the pMN domain at E12.5 and subsequently migrate to populate the entire neural tube. The development of these cells depends on SHH-mediated signalling and is inhibited by dorsally derived BMPs and WNT proteins. In the second wave (OPC 2, blue arrows), dorsal OPCs develop from *Olig2*-expressing cells of the dP3, dP4, dP5 and dP6 domains during the fetal phase (at day E15.5) in an SHH-independent manner. The origins of the third wave (OPC 3, red arrows), which occurs after birth, remain unclear. These OPCs might arise from progenitor cells, which remain around the central canal, or from proliferative *Ng2*-expressing precursor cells throughout the parenchyma. OPCs from ventral and dorsal regions are intermixed in the spinal cord at birth, with a heavy predominance of pMN-derived cells. The relative contributions at adult stages are unclear.

graded expression of transcription factors along the rostral–caudal and dorsal–ventral axes, resulting in the parcellation of the neuroepithelium into progenitor domains that lead to spatially distinct origins for most types of neuron[23].

Like in the spinal cord, patterning molecules establish spatial domains of proneural and homeodomain protein expression within forebrain progenitor cells. Proneural proteins are bHLH transcription factors — for example, ASCL1 (also known as MASH1), neurogenin 1 (NGN1), NGN2 and NGN3, and ATOH1 (also known as MATH1) — that initiate neurogenesis (reviewed in ref. 24). The commitment of progenitors to a neuronal fate involves not only the promotion of neurogenesis but also the coordinate inhibition of self-renewal and the suppression of gliogenic programs[25,26].

The switch from neuron to glial–cell production

Gliogenesis generally follows neurogenesis in the developing mammalian CNS, with the same progenitor domains switching developmental programs from neuron production mainly to oligodendrocyte or astrocyte production. In the ventral spinal cord, for example, there is a developmental switch from the production of neurons to oligodendrocytes within the same progenitor domain. (It should be kept in mind that radial glia in all domains of the CNS are thought to transform into astrocytes, which would yield a small proportion of astrocytes and ependymal cells, even in domains that primarily produce oligodendrocytes, such as the pMN domain[27].)

The spinal–cord pMN domain

The pMN domain of the ventral spinal cord, with its well-defined outputs of motor neurons and oligodendrocytes, is a suitable model for studying how the generation of successive waves of neuron and glial-cell progeny from a discrete precursor population is regulated. The actions of SHH and OLIGs are consistently required throughout the processes of pattern formation and motor neuron and oligodendrocyte specification, but several lines of evidence indicate that the 'tone' of both SHH-mediated signalling and the levels of OLIG2 proteins are crucial

for maintaining cell fate and the decision to produce differentiated first-wave motor neurons or, later, glia[28]. However, as described below, many other factors must be taken into account.

Proneural factors are repressors of gliogenesis. NGN2 is expressed in a subset of OLIG2-expressing cells in the pMN domain at the time of motor neuron production, and this pattern is conserved across species[18,19,29]. OPC production is preceded by downregulation of *Ngn2* (ref. 28), leading to the proposal that *Ngn2* downregulation is a determinant of the neuron–glial-cell switch. NGN activity and the neuron–glial-cell switch are also modulated in the pMN domain by factors involved in pattern formation. In contrast to early embryonic expression of NGN2 in the ventral neural tube, ASCL1 is expressed from E16 into the gliogenic phases, where it is required for oligodendrocyte development[30].

NOTCH proteins are also involved in gliogenesis. These proteins are single-pass, heterodimeric transmembrane receptors that bind to their transmembrane ligands (for example, the Delta-like proteins) at the surface of adjacent cells. Ligand binding results in cleavage of the intracellular domain of the receptor, which then translocates to the nucleus and recruits a complex that includes the transcriptional effector RBP-J. Forced expression of activated NOTCH1 promotes the formation of radial glia in the brain[31]. Analyses in zebrafish[29] and mice[32] show that oligodendrocytes fail to form in embryos that lack NOTCH signalling. Progenitors in the pMN domain contribute solely to motor neuron production in NOTCH mutants, whereas forced expression of activated NOTCH1 blocks neurogenesis and results in excess OPCs[33]. These studies indicate that the general role of NOTCH signalling is permissive rather than instructive for glial-cell fate acquisition.

'Pro-glial-cell' transcription factors are also required for the neuron–glial-cell switch. The studies discussed in the previous section are consistent with a broad range of data that illustrate the importance of the proneural protein–NOTCH program in regulating the decision to produce neurons or glia in the brain. However, additional 'pro-glial-cell' transcription programs are required.

Analysis of *Sox9* loss-of-function mutations in mice provided evidence that a gliogenic stage-specific transcriptional program is needed

for the generation of oligodendrocytes and astrocytes in the vertebrate neural tube[34]. *Sox9* mutants showed defects in the specification of oligodendrocytes in the pMN domain, as well as in the production of an apparent deficit of astrocytes in the p2 domain. Such animals also had more motor neurons and V2 interneurons, indicating that SOX9 is a general molecular component of the neuron–glial-cell switch in the developing spinal cord. Recent studies indicate that SOX9 also has a role in hindbrain cell-fate choice[35].

The initiation of gliogenesis in the embryonic spinal cord, and the differentiation of astrocytes later in gliogenesis, is controlled by the nuclear factor I (NFI) genes, which encode a family of transcription factors that bind CAATT boxes. Gain-of-function manipulations in the embryonic chick spinal cord indicate that NFIA expression is sufficient for gliogenesis, whereas loss of NFIA expression led to a loss of glial progenitors and a concomitant increase in neurogenesis that resulted from loss of NOTCH activity[36]. NFI proteins seems to have an instructive role in collaborating with NOTCH to promote gliogenesis. Furthermore, knockout of *Nfia* or *Nfib* results in decreased GFAP expression[37,38], in keeping with studies indicating that NFI genes directly regulate expression of this gene[39,40].

In summary, studies of the pMN domain indicate that the neuron–glial-cell switch is regulated in a complex manner that requires the following: ongoing (perhaps modulated) activity of SHH and OLIG2, downregulation of *Ngn2* expression, Delta-like–NOTCH signalling to preserve progenitors for producing second-wave domain-specific progeny, and activation of a pro-gliogenic phase-specific transcriptional program involving SOX9 and NFI transcription factors.

The neuron to glial–cell switch in the forebrain

Progenitor domains generate different types of neuron in a temporally specific manner. For example, in the dorsal pallium, laminar-specific neuron subtypes that are thought to arise from common progenitor cells are sequentially produced (deep-layer neurons first and upper-layer neurons later). This developmental program is largely cell-intrinsic, because cultured neural progenitors derived from the embryonic CNS, as well as embryonic-stem-cell-derived neural progenitors, generate neuron subtypes in a precise temporal and laminar order[41–43] but can also be affected by signals from the local environment[44]. In the anterior entopeduncular area, the homeobox proteins DLX1 and DLX2 regulate interneuron versus oligodendrocyte cell fate[45].

Early in forebrain development, the promoters of *GFAP* and *S100B* are methylated, and astrocyte development is repressed. At these early stages, neuroepithelial cells are insensitive to cytokines[46,47]. However, radial glia become competent to respond through demethylation of the promoters of astrocytic genes, possibly through epigenetic regulation[48]. When multipotent progenitor cells become competent to generate astrocytes, they are held in check by extrinsic signals that repress gliogenesis. For example, stimulation with FGF2 prompts the nuclear receptor co-repressor (NCOR) to act directly to the proximal *Gfap* promoter, thereby repressing transcription[49]. Neurogenesis is also promoted by neuregulin 1, which activates a presenilin-dependent nuclear signalling pathway involving ERBB4, a member of the epidermal growth-factor receptor family. This pathway antagonizes the actions of astrogenesis-promoting signals such as ciliary neurotrophic factor (CNTF)[50] through a mechanism that involves NCOR[49]. Interestingly, a developmental reduction in ERBB2 expression occurs at the end of neurogenesis when radial glia are transformed into astrocytes[51], and introduction of ERBB2 into adult astrocytes can restore their neurogenic potential[52]. Recent evidence highlights the role of epigenetic events in neuron–glial-cell fate decisions by inducing gene silencing through methylation and deacetylation. For example, the Polycomb group complex epigenetically suppresses the genes encoding proneural bHLH factors, promoting the neuron–astrocyte fate switch in neural precursor cells[53].

NOTCH signalling has been implicated in the switch to astrogenesis[51]. Neuron-committed intermediate progenitors and young neurons express the NOTCH ligands jagged 1 and Delta-like 1 (refs 54–56), which activate NOTCH signalling in radial glia, promoting astrogenesis. NOTCH promotes astrogenesis, in part, through HES proteins, which inhibit neurogenic bHLH factors, but also through promoting cytokine-mediated activation of the JAK–STAT pathway, leading to demethylation and upregulation of astrocyte-specific genes[57]. Thus, by increasing NOTCH activation, newly formed neurons may prime multi-potent progenitors to respond to gliogenic cytokines[54].

Gliogenesis might also be regulated by another neuronal feedback mechanism: cytokines secreted by neurons have been proposed to promote gliogenesis[58]. Neurons secrete gliogenic cytokines — particularly members of the interleukin 6 (IL-6) family, including leukaemia inhibitory factor (LIF), CNTF and cardiotrophin 1 (CT1; also known as CTF1) — which bind to a receptor complex that contains the α-subunit of the LIF receptor (LIFR) and gp130, activating the gp130–JAK–STAT pathway in cortical precursor cells and promoting gliogenesis[58–60]. Thus, mice lacking LIFR or gp130 have deficits in astrogenesis[61–63], and knockdown of expression of gp130 in cortical precursors decreases astrogenesis *in vitro* and *in vivo*[58]. During development of the cortex, CT1, in particular, has been shown to promote the neuron–glial-cell switch in multipotent precursor cells *in vivo*, as well as *in vitro*[58], supporting the concept that neuronal feedback may help to regulate the developmental switch to gliogenesis.

BMPs have a dual role in this switch, depending on the levels of certain growth factors that are present. BMPs promote neurogenesis during the neurogenic period and astrogenesis during the gliogenic period[64]. Exposure to BMP2 together with gliogenic cytokines promotes the formation of a SMAD–p300–CREB-binding-protein–STAT complex that transactivates astrocyte-associated genes[63]. At the same time, BMPs also exert a coordinate antagonism of proneural bHLH factors, suppressing neurogenesis[65] and perhaps gliogenesis. BMPs suppress oligodendrocyte development and promote astrocyte development. In addition, the proneural bHLH factor ASCL1 has a role in supporting oligodendrocyte development in the telencephalon in collaboration with OLIG2 (ref. 24) and an ongoing function in neuron and oligodendrocyte production in postnatal brain[66].

In contrast to the roles of OLIG2 in repressing astrogenesis in the embryonic spinal cord (described above), OLIG2 has also been proposed to support the determination of astrocyte cell fate in the forebrain. At day 7 after birth, in the subventricular zone of the mouse brain, OLIG2 was observed in gliogenic progenitors that produce both astrocytes and oligodendrocytes, and forced expression studies indicated that OLIG2 repressed differentiation into neurons and promoted glial-cell fates[67]. However, conditional ablation of *Olig2* in astrocytes expressing GFAP, and their precursors, reduced the proliferation of reactive astrocytes, suggesting that the functions of OLIG2 in astrocytes might relate to roles in cell cycle progression.

Temporally distinct waves of gliogenesis

So far we have focused on the initial phases of glial-cell specification. Now we turn to the later stages of gliogenesis, which differ in many respects, at least for oligodendrocytes.

Oligodendrocytes are produced in several distinct spatiotemporal waves in the spinal cord and brain

Although OPCs are derived from the ventral pMN domain in the embryonic spinal cord, there are additional sources of OPCs that emerge during fetal development in dorsal CNS[16,68]. Such fetal OPCs had a molecular phenotype distinct from their embryonic counterparts, with transient expression of the dorsal progenitor markers PAX7 and GSX1 and GSX2, but no functional differences between these cells have been demonstrated. Using a fate-mapping approach, Fogarty and colleagues[68] provided an additional line of evidence that OPCs develop from the ventricular-zone progenitors of the dorsal spinal cord. Together, these studies suggest that the contribution to the early postnatal spinal cord of ventral OPCs to dorsal OPCs is about 4:1. However, it remains to be determined whether the relative contributions of dorsal OPCs increase

with age, which would further clarify the functional importance of these cells.

In the ventral domains of both the spinal cord and the forebrain, SHH is required for OPC production[14,69]. Additional OPCs arise in an SHH-independent manner in the dorsal domains of the forebrain at later stages of development[69]. The precise origins of forebrain OPCs have been explored using Cre–loxP fate mapping in transgenic mice[70]. Beginning at about E11.5, OPCs were found in the medial ganglionic eminence and the anterior entopeduncular area. By E14.5, they were found throughout the telencephalon, including the cortex. Over time, intermediate and dorsal OPCs were produced and were observed to replace or dilute out early-wave OPCs. Indeed, the early-appearing OPCs derived from the medial ganglionic eminence and anterior entopeduncular area had mostly disappeared at 10 days after birth. Thus, three waves of OPCs arise to populate the forebrain, and these follow a ventral–dorsal temporal progression (Fig. 3).

The authors of this study[70] also used a Cre–loxP strategy to ablate early-wave OPCs specifically, by eliminating cells expressing a *Sox10* transgene through expression of a diphtheria-toxin-encoding transgene. Although there were no significant neurological consequences, further studies are needed to understand the precise developmental, phenotypic and functional differences between embryonic and fetal OPCs. These findings could have important implications in the setting of CNS injury. For example, distinct populations of OPCs might preferentially contribute to remyelination and/or be better targets for therapeutic manipulation.

Spatial and temporal patterns of astrogenesis

Thymidine-labelling experiments that detect proliferating cells indicated that, whereas cortical neurogenesis begins at about E12, oligodrogenesis begins around the time of birth. However, in these studies, the timing of glial-cell lineage specification was probably misjudged, because cells were labelled after exit from the cell cycle. A pool of committed precursors that proliferate early and differentiate later would therefore escape labelling. Indeed, as discussed above, specification of forebrain oligodendrocytes takes place in the embryo. Astrogenesis in the embryo is linked to the terminal phases of radial glial-cell function (Fig. 1). The precise timing remains unclear, however, because of a lack of definitive markers for astrocyte precursors (as distinct from radial glia) that could be used to monitor their emergence and proliferation at early developmental stages. In contrast to OPCs, which proliferate in the mammalian brain throughout life, mature astrocytes are generally thought to be quiescent but proliferate after injury[71]. However, further work is needed to determine the time frame of astrocyte precursor proliferation.

After astrocyte specification has occurred, astrocyte precursors migrate to their final positions, where they begin the process of terminal differentiation. Understanding the nature and extent of astrocyte precursor migration is essential for defining the relationship of mature astrocytes, if any, to their position of origin in the neural tube.

It is controversial whether astrocytes of the embryonic CNS migrate radially, tangentially or both. Most studies have relied on transplantation of cultured astrocytes or cortical explants into perinatal animals *in vivo* or in cortical slice cultures *in vitro*. These studies generally agree that astrocytes first migrate tangentially along white-matter tracts and then delaminate and move in a radial direction in the grey matter[72,73]. Some regions of the brain seem to exert a strong chemoattractive effect on astrocytes; for example, the substantia nigra attracts all astrocytes that have been transplanted in the vicinity of the midbrain[74]. However, a caveat of these studies is that *ex vivo* culture and transplantation conditions might select for an astrocyte subtype that is not representative of the type present in the embryo. In other studies, sparse retroviral labelling of proliferating multipotent precursors indicates that most clonally related cells (including astrocytes) are found in radial columns, as well as tangentially along fibre tracts[75]. It is possible that the modes of astrocyte migration vary in different regions of the nervous

system. Comprehensive *in vivo* analysis of astrocyte migration — using improved markers that have become available, imaging tools (for example, time-lapse microscopy) and sophisticated fate-mapping techniques (Box 1) — will be useful for answering these questions.

Future directions and challenges

Astrocytes are a heterogeneous group of cells, both functionally and morphologically. In recent years, researchers have made impressive progress in understanding which cell-intrinsic factors regulate OPC maturation. This has, however, highlighted a large gap in the knowledge of many fundamental aspects of astrocyte developmental biology.

Analysis of vertebrate embryos has improved the understanding of the mechanisms that underlie the specification of macroglia. As a departure from the classical models of glial-cell development through a bipotent glial-cell-restricted precursor, recent studies indicate that oligodendrocyte and astrocyte precursors *in vivo* develop separately and in mutually exclusive domains of the ventral neural tube. Broader implications of this segmental model of gliogenesis, however, remain to be defined. For example, could the region-restricted production of astrocytes help to establish the molecular and functional diversity of this population? In the case of the adult subventricular zone, one study indicates that the dorsal–ventral patterning of type B cells is reflected in permanent restrictions on their potential to produce different types of interneuron[76].

To answer such questions, regulatory factors that are specific for astrocytes and their precursors will probably need to be identified, and new genetic tools for exploring the complex functions of glia will be needed (Box 1). In addition, the role of invertebrate model systems in uncovering new pathways of gliogenesis cannot be underestimated. Indeed, researchers are making rapid progress using such approaches, which form an exciting area of research[77].

It is clear that macroglia are crucial for maintaining neurological function, as has been shown in studies of human diseases. In individuals with Pelizaeus–Merzbacher disease, a congenital leukodystrophy, mutation of *PLP1* renders OPCs defective and incapable of myelin production. An understanding of OPC development has contributed to the rationale for cell-based therapies for this disorder[78]. Oligodendrocytes are also targets in individuals with the autoimmune disease multiple sclerosis and in newborns with injuries to the white matter that are associated with cerebral palsy. Mutation of *GFAP* in humans is the aetiological factor in the congenital disorder Alexander disease, which typically has an adult onset. The list of astrocyte-based diseases is expanding to include amyotrophic lateral sclerosis[79], epilepsy[80] and Parkinson's disease[81]. The availability of robust and specific markers for the developmental progression of the glial-cell lineage will facilitate the assessment of the contributions of glia and their precursors to a range of human neurological disorders. For example, in multiple sclerosis and periventricular leukomalacia, markers (such as OLIG2) identified in developmental studies have been examined in neuropathological studies, showing that OPCs are blocked in their differentiation and fail to carry out normal repair functions after white-matter injury[82,83]. These findings suggest that regulation — rather than replacement — of endogenous progenitors is a promising therapeutic approach. Further progress in studying the developmental genetics of macroglia will undoubtedly improve our understanding of a range of human neurological disorders. ∎

1. Zalc, B. & Colman, D. R. Origins of vertebrate success. *Science* **288,** 271–272 (2000).
2. Roots, B. I. The phylogeny of invertebrates and the evolution of myelin. *Neuron Glia Biol.* **4,** 101–109 (2008).
3. Lin, S. C. & Bergles, D. E. Synaptic signaling between GABAergic interneurons and oligodendrocyte precursor cells in the hippocampus. *Nature Neurosci.* **7,** 24–32 (2004).
   This paper shows that OPCs of the hippocampus receive direct input from interneurons, suggesting that GABA (γ-aminobutyric acid)-induced chloride flux influences oligodendrocyte development and the efficacy of glutamate-induced signalling in OPCs.
4. Malatesta, P., Hartfuss, E. & Gotz, M. Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage.

Development **127**, 5253–5263 (2000).
This paper highlights the neurogenic potential of radial glia. Using fluorescence-activated cell sorting, radial glia isolated from the embryonic neocortex were shown to generate neurons, as well as astrocytes.

5. Miyata, T., Kawaguchi, A., Okano, H. & Ogawa, M. Asymmetric inheritance of radial glial fibers by cortical neurons. *Neuron* **31**, 727–741 (2001).

6. Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).
This study used retroviral labelling *in utero* and showed that neurons were directly generated through the self-renewal division of radial glia.

7. Anthony, T. E., Klein, C., Fishell, G. & Heintz, N. Radial glia serve as neuronal progenitors in all regions of the central nervous system. *Neuron* **41**, 881–890 (2004).

8. Haubensak, W., Attardo, A., Denk, W. & Huttner, W. B. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc. Natl Acad. Sci. USA* **101**, 3196–3201 (2004).

9. Miyata, T. *et al.* Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development* **131**, 3133–3145 (2004).

10. Noctor, S. C., Martinez-Cerdeno, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nature Neurosci.* **7**, 136–144 (2004).

11. Noctor, S. C., Martinez-Cerdeno, V. & Kriegstein, A. R. Distinct behaviors of neural stem and progenitor cells underlie cortical neurogenesis. *J. Comp. Neurol.* **508**, 28–44 (2008).

12. Briscoe, J. & Novitch, B. G. Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. *Phil. Trans. R. Soc. B* **363**, 57–70 (2008).

13. Ulloa, F. & Marti, E. Wnt won the war: antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev. Dyn.* **239**, 69–76 (2010).

14. Orentas, D. M., Hayes, J. E., Dyer, K. L. & Miller, R. H. Sonic hedgehog signaling is required during the appearance of spinal cord oligodendrocyte precursors. *Development* **126**, 2419–2429 (1999).

15. Dessaud, E. *et al.* Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* **450**, 717–720 (2007).

16. Vallstedt, A., Klos, J. M. & Ericson, J. Multiple dorsoventral origins of oligodendrocyte generation in the spinal cord and hindbrain. *Neuron* **45**, 55–67 (2005).

17. Lu, Q. R. *et al.* Common developmental requirement for *Olig* function indicates a motor neuron/oligodendrocyte connection. *Cell* **109**, 75–86 (2002).

18. Zhou, Q. & Anderson, D. J. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* **109**, 61–73 (2002).

19. Novitch, B. G., Chen, A. I. & Jessell, T. M. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. *Neuron* **31**, 773–789 (2001).

20. Muroyama, Y., Fujiwara, Y., Orkin, S. H. & Rowitch, D. H. Specification of astrocytes by bHLH protein SCL in a restricted region of the neural tube. *Nature* **438**, 360–363 (2005).
This paper was the first to suggest that astrocyte development is regulated by region-restricted mechanisms.

21. Hochstim, C., Deneen, B., Lukaszewicz, A., Zhou, Q. & Anderson, D. J. Identification of positionally distinct astrocyte subtypes whose identities are specified by a homeodomain code. *Cell* **133**, 510–522 (2008).
This paper shows that a homeodomain code in the ventral neural tube regulates the diversity of astrocytes. It provides the strongest evidence so far for the 'segmental model' of astrocyte development.

22. Genethliou, N. *et al.* Spatially distinct functions of PAX6 and NKX2.2 during gliogenesis in the ventral spinal cord. *Biochem. Biophys. Res. Commun.* **382**, 69–73 (2009).

23. Hoch, R. V., Rubenstein, J. L. & Pleasure, S. Genes and signaling events that establish regional patterning of the mammalian forebrain. *Semin. Cell Dev. Biol.* **20**, 378–386 (2009).

24. Parras, C. M. *et al.* The proneural gene *Mash1* specifies an early population of telencephalic oligodendrocytes. *J. Neurosci.* **27**, 4233–4242 (2007).

25. Bylund, M., Andersson, E., Novitch, B. G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nature Neurosci.* **6**, 1162–1168 (2003).

26. Sandberg, M., Kallstrom, M. & Muhr, J. Sox21 promotes the progression of vertebrate neurogenesis. *Nature Neurosci.* **8**, 995–1001 (2005).

27. Masahira, N. *et al.* Olig2-positive progenitors in the embryonic spinal cord give rise not only to motoneurons and oligodendrocytes, but also to a subset of astrocytes and ependymal cells. *Dev. Biol.* **293**, 358–369 (2006).

28. Lee, S. K., Lee, B., Ruiz, E. C. & Pfaff, S. L. Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. *Genes Dev.* **19**, 282–294 (2005).

29. Park, H. C. & Appel, B. Delta–Notch signaling regulates oligodendrocyte specification. *Development* **130**, 3747–3755 (2003).

30. Battiste, J. *et al.* Ascl1 defines sequentially generated lineage-restricted neuronal and oligodendrocyte precursor cells in the spinal cord. *Development* **134**, 285–293 (2007).

31. Gaiano, N., Nye, J. S. & Fishell, G. Radial glial identity is promoted by Notch1 signaling in the murine forebrain. *Neuron* **26**, 395–404 (2000).

32. Yang, X. *et al.* Notch1 signaling influences v2 interneuron and motor neuron development in the spinal cord. *Dev. Neurosci.* **28**, 102–117 (2006).

33. Itoh, M. *et al.* Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta. *Dev. Cell* **4**, 67–82 (2003).

34. Karcavich, R. & Doe, C. Q. *Drosophila* neuroblast 7-3 cell lineage: a model system for studying programmed cell death, Notch/Numb signaling, and sequential specification of ganglion mother cell identity. *J. Comp. Neurol.* **481**, 240–251 (2005).

35. Esain, V., Postlethwait, J. H., Charnay, P. & Ghislain, J. FGF-receptor signalling controls neural cell diversity in the zebrafish hindbrain by regulating *olig2* and *sox9*. *Development* **137**, 33–42 (2010).

36. Deneen, B. *et al.* The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* **52**, 953–968 (2006).
In this paper, the 'pro-glial-cell' transcription factor NFIA is described and shown to promote gliogenesis, inhibit neurogenesis in ventricular-zone spinal-cord progenitors and, at later stages, promote differentiation into astrocytes.

37. das Neves, L. *et al.* Disruption of the murine nuclear factor I-A gene (*Nfia*) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. *Proc. Natl Acad. Sci. USA* **96**, 11946–11951 (1999).

38. Steele-Perkins, G. *et al.* The transcription factor gene *Nfib* is essential for both lung maturation and brain development. *Mol. Cell. Biol.* **25**, 685–698 (2005).

39. Cebolla, B. & Vallejo, M. Nuclear factor-I regulates glial fibrillary acidic protein gene expression in astrocytes differentiated from cortical precursor cells. *J. Neurochem.* **97**, 1057–1070 (2006).

40. Wilczynska, K. M. *et al.* Nuclear factor I isoforms regulate gene expression during the differentiation of human neural progenitors to astrocytes. *Stem Cells* **27**, 1173–1181 (2009).

41. Gaspard, N. *et al.* An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* **455**, 351–357 (2008).

42. Shen, Q. *et al.* The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. *Nature Neurosci.* **9**, 743–751 (2006).

43. Eiraku, M. *et al.* Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532 (2008).

44. Desai, A. R. & McConnell, S. K. Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development* **127**, 2863–2872 (2000).

45. Petryniak, M. A., Potter, G. B., Rowitch, D. H. & Rubenstein, J. L. Dlx1 and Dlx2 control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron* **55**, 417–433 (2007).
This paper shows that DLX1 and DLX2 repress oligodendrocyte production in an area of the forebrain and suggests that DLX1 and DLX2 have non-cell-autonomous effects or epigenetic effects in the oligodendrocyte lineage.

46. Temple, S. The development of neural stem cells. *Nature* **414**, 112–117 (2001).

47. Molne, M. *et al.* Early cortical precursors do not undergo LIF-mediated astrocytic differentiation. *J. Neurosci. Res.* **59**, 301–311 (2000).

48. Takizawa, T. *et al.* DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. *Dev. Cell* **1**, 749–758 (2001).

49. Hermanson, O., Jepsen, K. & Rosenfeld, M. G. N-CoR controls differentiation of neural stem cells into astrocytes. *Nature* **419**, 934–939 (2002).

50. Sardi, S. P., Murtie, J., Koirala, S., Patten, B. A. & Corfas, G. Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain. *Cell* **127**, 185–197 (2006).

51. Fox, I. J. & Kornblum, H. I. Developmental profile of ErbB receptors in murine central nervous system: implications for functional interactions. *J. Neurosci. Res.* **79**, 584–597 (2005).

52. Ghashghaei, H. T. *et al.* Reinduction of ErbB2 in astrocytes promotes radial glial progenitor identity in adult cerebral cortex. *Genes Dev.* **21**, 3258–3271 (2007).

53. Hirabayashi, Y. *et al.* Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition. *Neuron* **63**, 600–613 (2009).

54. Namihira, M. *et al.* Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. *Dev. Cell* **16**, 245–255 (2009).

55. Campos, L. S., Duarte, A. J., Branco, T. & Henrique, D. mDll1 and mDll3 expression in the developing mouse brain: role in the establishment of the early cortex. *J. Neurosci. Res.* **64**, 590–598 (2001).

56. Yoon, K. J. *et al.* Mind bomb 1-expressing intermediate progenitors generate Notch signaling to maintain radial glial cells. *Neuron* **58**, 519–531 (2008).

57. Kamakura, S. *et al.* Hes binding to STAT3 mediates crosstalk between Notch and JAK–STAT signalling. *Nature Cell Biol.* **6**, 547–554 (2004).

58. Barnabe-Heider, F. *et al.* Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1. *Neuron* **48**, 253–265 (2005).
This paper provides evidence that embryonic cortical neurons regulate astrocyte development by secreting the neurotrophic cytokine CT1, which activates the gp130–JAK–STAT signalling pathway.

59. Bonni, A. *et al.* Regulation of gliogenesis in the central nervous system by the JAK–STAT signaling pathway. *Science* **278**, 477–483 (1997).

60. Ochiai, W., Yanagisawa, M., Takizawa, T., Nakashima, K. & Taga, T. Astrocyte differentiation of fetal neuroepithelial cells involving cardiotrophin-1-induced activation of STAT3. *Cytokine* **14**, 264–271 (2001).

61. Ware, C. B. *et al.* Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes placental, skeletal, neural and metabolic defects and results in perinatal death. *Development* **121**, 1283–1299 (1995).

62. Koblar, S. A. *et al.* Neural precursor differentiation into astrocytes requires signaling through the leukemia inhibitory factor receptor. *Proc. Natl Acad. Sci. USA* **95**, 3178–3181 (1998).

63. Nakashima, K. *et al.* Synergistic signaling in fetal brain by STAT3–Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).

64. Mabie, P. C., Mehler, M. F. & Kessler, J. A. Multiple roles of bone morphogenetic

protein signaling in the regulation of cortical cell number and phenotype. *J. Neurosci.* **19,** 7077–7088 (1999).

65. Nakashima, K. *et al.* BMP2-mediated alteration in the developmental pathway of fetal mouse brain cells from neurogenesis to astrocytogenesis. *Proc. Natl Acad. Sci. USA* **98,** 5868–5873 (2001).

66. Parras, C. M. *et al.* Mash1 specifies neurons and oligodendrocytes in the postnatal brain. *EMBO J.* **23,** 4495–4505 (2004).

67. Marshall, C. A., Novitch, B. G. & Goldman, J. E. Olig2 directs astrocyte and oligodendrocyte formation in postnatal subventricular zone cells. *J. Neurosci.* **25,** 7289–7298 (2005).

68. Fogarty, M., Richardson, W. D. & Kessaris, N. A subset of oligodendrocytes generated from radial glia in the dorsal spinal cord. *Development* **132,** 1951–1959 (2005).

69. Nery, S., Wichterle, H. & Fishell, G. Sonic hedgehog contributes to oligodendrocyte specification in the mammalian forebrain. *Development* **128,** 527–540 (2001).

70. Kessaris, N. *et al.* Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nature Neurosci.* **9,** 173–179 (2006).
    This study used Cre–*loxP* fate-mapping technology and showed that oligodendrocytes develop in at least three distinct waves from the time of embryogenesis through the postnatal stages.

71. Buffo, A. *et al.* Origin and progeny of reactive gliosis: a source of multipotent cells in the injured brain. *Proc. Natl Acad. Sci. USA* **105,** 3581–3586 (2008).

72. Hatton, J. D., Nguyen, M. H. & U, H. S. Differential migration of astrocytes grafted into the developing rat brain. *Glia* **9,** 113–119 (1993).

73. Jacobsen, C. T. & Miller, R. H. Control of astrocyte migration in the developing cerebral cortex. *Dev. Neurosci.* **25,** 207–216 (2003).

74. Zhou, H. F. & Lund, R. D. Migration of astrocytes transplanted to the midbrain of neonatal rats. *J. Comp. Neurol.* **317,** 145–155 (1992).

75. Gray, G. E. & Sanes, J. R. Migratory paths and phenotypic choices of clonally related cells in the avian optic tectum. *Neuron* **6,** 211–225 (1991).

76. Merkle, F. T., Mirzadeh, Z. & Alvarez-Buylla, A. Mosaic organization of neural stem cells in the adult brain. *Science* **317,** 381–384 (2007).
    This paper shows that in the adult the subventricular zone contains regions that produce different types of interneuron, which contribute to the rostral migratory stream and the olfactory bulb.

77. Freeman, M. Specification and morphogenesis of astrocytes. *Science* (in the press).

78. Goldman, S. A., Schanz, S. & Windrem, M. S. Stem cell-based strategies for treating pediatric disorders of myelin. *Hum. Mol. Genet.* **17,** R76–R83 (2008).

79. Nagai, M. *et al.* Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. *Nature Neurosci.* **10,** 615–622 (2007).

80. Nicoll, J. A. & Weller, R. O. A new role for astrocytes: β-amyloid homeostasis and degradation. *Trends Mol. Med.* **9,** 281–282 (2003).

81. Saijo, K. *et al.* A Nurr1/CoREST pathway in microglia and astrocytes protects dopaminergic neurons from inflammation-induced death. *Cell* **137,** 47–59 (2009).

82. Chang, A., Tourtellotte, W. W., Rudick, R. & Trapp, B. D. Premyelinating oligodendrocytes in chronic lesions of multiple sclerosis. *N. Engl. J. Med.* **346,** 165–173 (2002).

83. Billiards, S. S. *et al.* Myelin abnormalities without oligodendrocyte loss in periventricular leukomalacia. *Brain Pathol.* **18,** 153–163 (2008).

84. Richardson, W. D., Kessaris, N. & Pringle, N. Oligodendrocyte wars. *Nature Rev. Neurosci.* **7,** 11–18 (2006).

85. Raff, M. C., Miller, R. H. & Noble, M. A glial progenitor cell that develops *in vitro* into an astrocyte or an oligodendrocyte depending on culture medium. *Nature* **303,** 390–396 (1983).

86. Herrera, J. *et al.* Embryonic-derived glial-restricted precursor cells (GRP cells) can differentiate into astrocytes and oligodendrocytes *in vivo. Exp. Neurol.* **171,** 11–21 (2001).

87. Rao, M. S., Noble, M. & Mayer-Proschel, M. A tripotential glial precursor cell is present in the developing spinal cord. *Proc. Natl Acad. Sci. USA* **95,** 3996–4001 (1998).

88. Costa, M. R., Bucholz, O., Schroeder, T. & Gotz, M. Late origin of glia-restricted progenitors in the developing mouse cerebral cortex. *Cereb. Cortex* **19** (suppl. 1), i135–i143 (2009).

89. Wu, S., Wu, Y. & Capecchi, M. R. Motoneurons and oligodendrocytes are sequentially generated from neural stem cells but do not appear to share common lineage-restricted progenitors *in vivo. Development* **133,** 581–590 (2006).

90. Levison, S. W. & Goldman, J. E. Both oligodendrocytes and astrocytes develop from progenitors in the subventricular zone of postnatal rat forebrain. *Neuron* **10,** 201–212 (1993).

91. Parnavelas, J. G. Glial cell lineages in the rat cerebral cortex. *Exp. Neurol.* **156,** 418–429 (1999).

92. Zhao, J. W., Raha-Chowdhury, R., Fawcett, J. W. & Watts, C. Astrocytes and oligodendrocytes can be generated from NG2⁺ progenitors after acute brain injury: intracellular localization of oligodendrocyte transcription factor 2 is associated with their fate choice. *Eur. J. Neurosci.* **29,** 1853–1869 (2009).

93. Tatsumi, K. *et al.* Genetic fate mapping of Olig2 progenitors in the injured adult cerebral cortex reveals preferential differentiation into astrocytes. *J. Neurosci. Res.* **86,** 3494–3502 (2008).

94. Menn, B. *et al.* Origin of oligodendrocytes in the subventricular zone of the adult brain. *J. Neurosci.* **26,** 7907–7918 (2006).

95. Jablonska, B. *et al.* Chordin-induced lineage plasticity of adult SVZ neuroblasts after demyelination. *Nature Neurosci.* **13,** 541–550 (2010).

96. Raymond, C. S. & Soriano, P. High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells. *PLoS ONE* **2,** e162 (2007).

97. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121,** 479–492 (2005).

98. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450,** 56–62 (2007).

99. Pringle, N. P. *et al. Fgfr3* expression by astrocytes and their precursors: evidence that astrocytes and oligodendrocytes originate in distinct neuroepithelial domains. *Development* **130,** 93–102 (2003).

100. Shibata, T. *et al.* Glutamate transporter GLAST is expressed in the radial glia–astrocyte lineage of developing mouse spinal cord. *J. Neurosci.* **17,** 9212–9219 (1997).

101. Owada, Y., Yoshimoto, T. & Kondo, H. Spatio-temporally differential expression of genes for three members of fatty acid binding proteins in developing and mature rat brains. *J. Chem. Neuroanat.* **12,** 113–122 (1996).

102. Stolt, C. C. *et al.* The Sox9 transcription factor determines glial fate choice in the developing spinal cord. *Genes Dev.* **17,** 1677–1689 (2003).

103. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28,** 264–278 (2008).

104. Fu, H. *et al.* A genome-wide screen for spatially restricted expression patterns identifies transcription factors that regulate glial development. *J. Neurosci.* **29,** 11399–11408 (2009).

105. Anthony, T. E. & Heintz, N. The folate metabolic enzyme ALDH1L1 is restricted to the midline of the early CNS, suggesting a role in human neural tube defects. *J. Comp. Neurol.* **500,** 368–383 (2007).

REVIEW

Regulation of synaptic connectivity by glia

Cagla Eroglu[1] & Ben A. Barres[2]

The human brain contains more than 100 trillion (10^{14}) synaptic connections, which form all of its neural circuits. Neuroscientists have long been interested in how this complex synaptic web is weaved during development and remodelled during learning and disease. Recent studies have uncovered that glial cells are important regulators of synaptic connectivity. These cells are far more active than was previously thought and are powerful controllers of synapse formation, function, plasticity and elimination, both in health and disease. Understanding how signalling between glia and neurons regulates synaptic development will offer new insight into how the nervous system works and provide new targets for the treatment of neurological diseases.

Glial cells occupy more than half of the volume of the human brain. There are several types in the central nervous system (CNS), including astrocytes, oligodendrocytes and microglia[1]. In addition, a new class of glial cell, oligodendrocyte precursor cells (OPCs) that express the proteoglycan NG2, has recently been identified and has morphological and physiological features that are distinct from those of other glia[2,3]. Astrocytes and oligodendrocyte lineage cells are derived from neural stem cells, whereas microglia originate from the immune system[4]. In the peripheral nervous system, there are two classes of Schwann cell (myelinating and non-myelinating), which functionally and antigenically resemble the glia of the CNS[5].

Glia are vital for the survival and function of neurons. Oligodendrocytes and Schwann cells myelinate axons to ensure fast, saltatory movement of action potentials. Astrocytes regulate blood flow, provide much-needed energy to neurons, and supply the building blocks of neurotransmitters, which fuel synapse function. But the functions of glia are not restricted to supporting neuronal function[4,6].

In this Review, we describe the numerous recent findings that illustrate the importance of glia in the formation, function, plasticity and elimination of synapses in the nervous system. We also discuss how these findings provide new insight into the pathophysiology of chronic pain, neurological diseases such as epilepsy, and neurodegenerative disorders such as Alzheimer's disease and glaucoma.

Glia are intimately associated with synapses

In the peripheral nervous system, synapses are ensheathed by non-myelinating Schwann cells, and in the CNS by astrocytes (Fig. 1a). The CNS also contains two forms of elongated, radial glial cell: Bergmann glia in the cerebellum, and Müller cells in the retina. These have many features in common with astrocytes and are closely associated with synapses[7]. Such an association of glia with synapses seems to have been conserved across evolution, because a class of synapse-ensheathing glial cell is also found in the brain of the fruitfly *Drosophila melanogaster* and has surprisingly similar morphology to rodent astrocytes[8]. This structural association extends to function. Perisynaptic glia ensure potassium ion homeostasis and regulate extracellular pH (Fig. 1b). Moreover, these cells express several receptors for neurotransmitters, enabling them to 'listen' to synapse function and respond to synapse activity by making localized and global changes in intracellular calcium ion concentrations[9,10] (Fig. 1b). In addition, glia modulate the properties of synapses by releasing neurologically active substances such as ATP and D-serine[4]. The extensive structural and functional association of perisynaptic glia

with the synapse gave rise to the concept of the 'tripartite synapse', in which synapses are defined as comprising the presynaptic and postsynaptic specializations of the neurons and the glial process that ensheaths them[11] (Fig. 1b).

In the past decade surprising new findings showed that one glial subtype receives synaptic inputs. Fast synaptic transmission occurs between OPCs and axons, both in the hippocampus and the cerebellum[12,13]. These OPCs can receive input mediated by the neurotransmitters glutamate and GABA (γ-aminobutyric acid)[14,15]. The functional significance of these neuron-to-glia synapses is not known. Although a matter of speculation at present, it is possible that neuronal activity regulates the differentiation of OPCs into myelinating oligodendrocytes according to the requirements of the neural circuit.

Astrocytes are the main population of glia in the brain. Rodent astrocytes have been classified into two groups on the basis of their morphology and location. One group contains the protoplasmic astrocytes of the grey matter, which are highly ramified. These cells ensheath synapses and are in contact with blood vessels. The other group is the fibrous astrocytes of the white matter, cells that are in contact with the nodes of Ranvier[1]. Recent physiological and gene expression profiling studies indicate that astrocytes, like neurons, are a diverse cell population, with distinct properties in different brain regions and at different periods of development. So this classification into two groups might not be adequate to appreciate the full extent of astrocyte diversity[16]. Moreover, the number and size of astrocytes in the brain varies between species and increases with species brain size and cognitive capability. For example, the human brain contains several more populations of astrocytes than the rodent brain, and human astrocytes are up to threefold larger and more ramified than their rodent counterparts[17].

The processes of protoplasmic astrocytes infiltrate into the neuropil and wrap themselves around synapses[18–20]. In the hippocampus, individual astrocytes parcel out the neuropil in a non-overlapping manner to form separate anatomical domains[18]. A similar tile-like organization of glia is characteristic of the anatomical distribution of cortical astrocytes[19] and also of cerebellar Bergmann glia, which ensheath most Purkinje cell synapses[21].

Astrocytes do not readily wrap around all synapses and the presence and extent of astrocyte coverage might be regulated[20]. In the hippocampus, protoplasmic astrocytes have been found to ensheath 57% of the synapses, most of which are excitatory synapses[22]. What directs astrocyte processes to synapses? It has been postulated that astrocytes extend processes towards regions where the glutamate concentration is

[1]Cell Biology Department, Box 3709, Duke University Medical Center, Durham, North Carolina 27710, USA. [2]Department of Neurobiology, Stanford University Medical Center, Fairchild Building D-200, Stanford, California 94305, USA.

higher. In agreement with this hypothesis, in the hippocampus of adult rats, large perforated synapses with higher probabilities of glutamate release were found to be more likely to be ensheathed by astrocytes[22]. It is equally possible, however, that synapses that are wrapped by an astrocyte are stabilized and receive nurturing signals that allow them to mature further. It is important to point out that astrocyte processes do not fully insulate a synapse from the surrounding environment but, instead, allow some flow into and out of the synaptic cleft. For example, at hippocampal synapses, regardless of their size, there are astrocyte-free perimeters, where substances could escape from or enter the synaptic clefts. These regions might be important for the *trans*-synaptic activation of glutamate receptors[20].

Association of synapses and astrocyte processes might be a dynamic process. In agreement with this possibility, astrocyte coverage can be altered during development in response to injury and in various physiological conditions[23–26]. Live imaging of organotypic hippocampal slices shows that astrocytes rapidly extend and retract their processes to engage and disengage from postsynaptic dendritic spines[27,28]. Astrocytes are more motile than their dendritic counterpart, and their movement is regulated by actin dynamics. Interestingly, dendritic protrusions that had contacts with astrocytes survived longer and were morphologically more mature than those without such contacts[28]. These observations suggest that astrocyte processes might control the stabilization of individual dendritic protrusions and their subsequent maturation into spines, and they present the possibility that astrocytes could have a crucial role in the experience-dependent structural synaptic changes that underlie learning and memory.

A series of studies in the adult rodent hippocampus suggested that the receptor tyrosine kinase EphA4 on the dendritic spines of pyramidal neurons regulates spine morphology by interacting with its ligand ephrin-A3, which is located in the perisynaptic processes of astrocytes[29]. Perturbing ephrin–EphA signalling led to increased numbers of glutamate transporters on astrocytes, aberrant morphology of dendritic spines, and defects in hippocampal learning[30] and long-term potentiation[31].

A remarkable interplay between astrocyte processes and synapses occurs in the hypothalamo-hypophyseal system at the hypothalamic supraoptic nucleus. In physiological conditions such as parturition, lactation and chronic dehydration, most of the astrocyte processes in contact with the soma and dendrites of magnocellular oxytocin neurons denervate these neurons in an oxytocin-dependent manner[32]. Concurrently, new inhibitory and excitatory synapses are formed onto these neurons, and these disappear when stimulation ends and the astrocyte processes reinnervate these oxytocinergic neurons[33].

Taken together, these findings show that different glial cell types regulate different aspects of nervous-system architecture, function and plasticity, through dynamic and often bidirectional structural interactions with synapses.

Glia control synapse formation

Establishment of the correct number and type of synapses is crucial for the proper development and function of the human brain. Many presynaptic and postsynaptic 'target-derived' factors have been implicated as important regulators of synapse formation and specificity[34]. Because astrocytes are an integral part of synapses, they probably contribute to the establishment of synapses. In addition, gliogenesis and synaptogenesis occur concurrently in the brain, and glial-cell maturation marks the end of the synaptogenic and plastic periods[35,36].

In vitro studies using purified neurons and astrocytes paved the way for understanding the extent to which astrocytes influence the shaping of synapse formation. They showed that synaptogenesis is not solely controlled by neurons but, instead, can be instructed by glia-derived signals. Retinal ganglion cells (RGCs), unlike many other CNS neurons, can be isolated from rodent retina as a purified population and can then be cultured in serum-free media of known composition[37]. Under these conditions, RGCs show little spontaneous synapse activity and
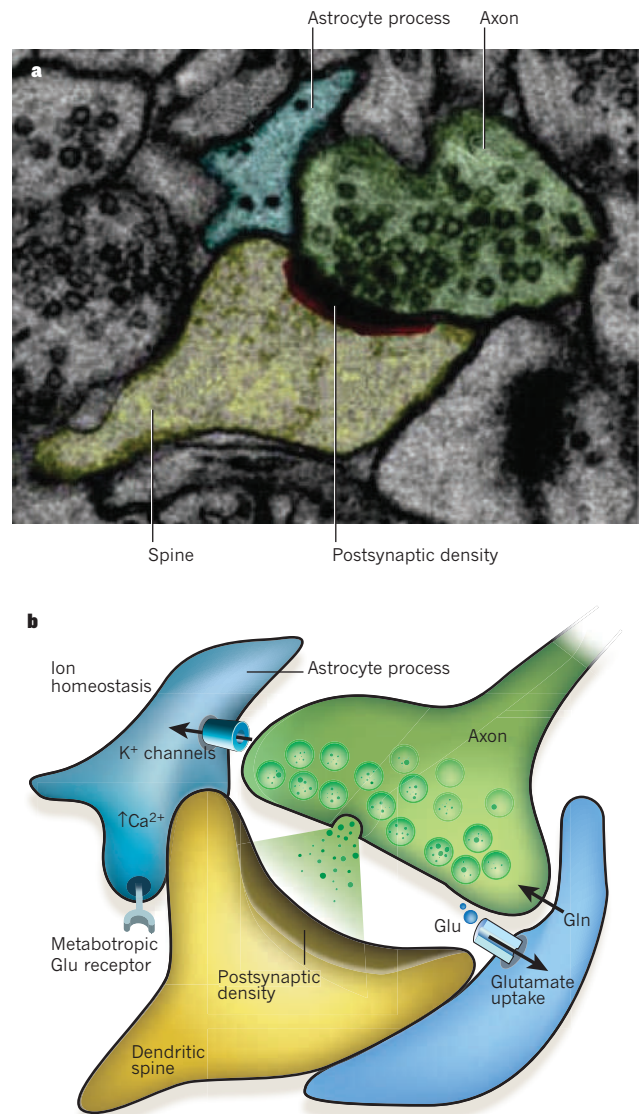
Figure 1 | The tripartite synapse. The processes of astrocytes are intimately associated with synapses. This association is both structural and functional. **a**, Electron micrograph showing a tripartite synapse in the hippocampus. The astrocyte process (blue) ensheaths the perisynaptic area. The axon of the neuron is shown in green, with the dendritic spine in yellow and the postsynaptic density in red and black. Reproduced, with permission, from ref. 22. **b**, Schematic representation of a tripartite synapse. Perisynaptic astrocyte processes contain transporters that take up glutamate (Glu, green circles) that has been released into the synapse and return it to neurons in the form of glutamine (Gln). Glutamate receptors on astrocytes (such as metabotropic glutamate receptors) sense synaptic glutamate release, which in turn induces a rise in Ca^{2+} concentration in the astrocytes. One of the main functions of glia at the synapse is to maintain ion homeostasis, for example regulating extracellular K^+ concentrations and pH.

form few synapses, whereas RGCs cultured in the presence of a feeding layer of astrocytes or astrocyte-conditioned medium show about tenfold more excitatory synapse activity and a fivefold to sevenfold increase in the number of synapses[38,39]. In addition to rat RGCs, astrocytes have been shown to be necessary for inducing synapse formation by human neurons generated from embryonic stem cells[40]. Astrocyte-induced synapses are ultrastructurally normal with functional excitatory synaptic events mediated by AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors[39]. Several studies using this system uncovered the presence of at least three classes of factor secreted by astrocytes: those that induce the formation of structurally normal but postsynaptically

silent synapses; those that facilitate presynaptic activity and enhance the probability of neurotransmitter release; and those that stimulate the insertion of glutamate receptors, which are stored in intracellular vesicles to the postsynaptic membranes, thus facilitating postsynaptic function[1] (Fig. 2).

The synapse-formation-inducing signals from astrocytes were identified to be a family of extracellular matrix proteins called thrombospondins (TSPs)[41]. Purified TSPs alone have been found to increase synapse number in RGC cultures to a level comparable to that induced by culturing in astrocyte-conditioned medium. TSP-induced synapses are ultrastructurally normal, presynaptically active but postsynaptically silent because of a lack of AMPA receptors on the postsynaptic side. Removal of TSPs from astrocyte-conditioned medium diminishes the synaptogenic activity of the medium. These results show that TSPs are the necessary and sufficient synaptogenic factors in astrocyte-conditioned medium for inducing the formation of structural synapses. *In vivo*, TSP1 and TSP2 are expressed by developing astrocytes at early postnatal stages, when most excitatory synapses are forming, and their expression is downregulated in adults. In addition, mice lacking both TSP1 and TSP2 have significantly fewer excitatory synapses in the cortex, showing that TSPs are important for excitatory synapse formation *in vivo*[41].

Using molecular and biochemical techniques, the $\alpha_2\delta$-1 subunit (encoded by the gene *CACNA2D1*) of the voltage-dependent Ca^{2+} channel complex was identified as the relevant neuronal receptor for TSPs[42]. All five mammalian TSPs can induce synapse formation by binding through their type 2 epidermal-growth-factor-like repeats to the von Willebrand factor type A (vWFA) domain of neuronal $\alpha_2\delta$-1. It is postulated that the interaction between TSPs and $\alpha_2\delta$-1 triggers cellular events that lead to synapse formation by activating a synaptogenic signalling complex, which might include Ca^{2+} channels[42]. TSPs also bind to integrins, which are important regulators of synapse structure and function[43,44]. Moreover, TSP1 was recently identified to be a ligand for the neuroligin family of synaptic adhesion proteins[45]. Given its ability to interact with multiple molecule types, it is possible that TSPs can not only induce synapse formation but can also modulate presynaptic and postsynaptic function. Understanding how TSP–$\alpha_2\delta$-1 signalling in neurons induces synapse formation has the potential to provide new insight into the molecular basis of this process, which has long been mysterious.

The $\alpha_2\delta$-1 subunit is also the receptor for the drug gabapentin (Neurontin), which is used to treat neuropathic pain and epilepsy and has an unknown mechanism of action[46]. Gabapentin blocks TSP-induced excitatory synaptogenesis *in vitro* and markedly inhibits excitatory synapse formation between neurons throughout the entire developing brain. It prevents this type of synapse formation by blocking the ability of TSPs to bind to $\alpha_2\delta$-1, thus inhibiting synaptogenic signalling initiated by the TSP–$\alpha_2\delta$-1 interaction without dissolving previously formed synapses[42]. These findings provide an additional line of evidence highlighting the ability of astrocytes as powerful promoters of synapse formation *in vivo*. They also suggest that TSP–$\alpha_2\delta$-1 signalling and astrocyte-induced synapse formation might be involved in the pathophysiology of disorders such as neuropathic pain and epilepsy.

As astrocytes can induce both presynaptic and postsynaptic activity, whereas TSPs can only instruct the formation of postsynaptically silent synapses, there must be other signals secreted by astrocytes that facilitate the insertion of glutamate receptors at the postsynaptic site and thus convert silent synapses to active ones. Through these signals, astrocytes can regulate synapse strength and plasticity. The identity of the astrocyte-secreted signal that can induce AMPA receptor insertion into synapses is unknown. Cholesterol in apolipoprotein E particles secreted by astrocytes, however, has been shown to increase the induced synaptic responses substantially in autaptic, cultured RGCs[47] by increasing presynaptic function and dendrite differentiation in this system[41,48]. The relevance of cholesterol and apolipoprotein E in synapse function *in vivo* is not yet known.

a Silent but structurally normal **b** Enhanced presynaptic function **c** Postsynaptically active (functional)
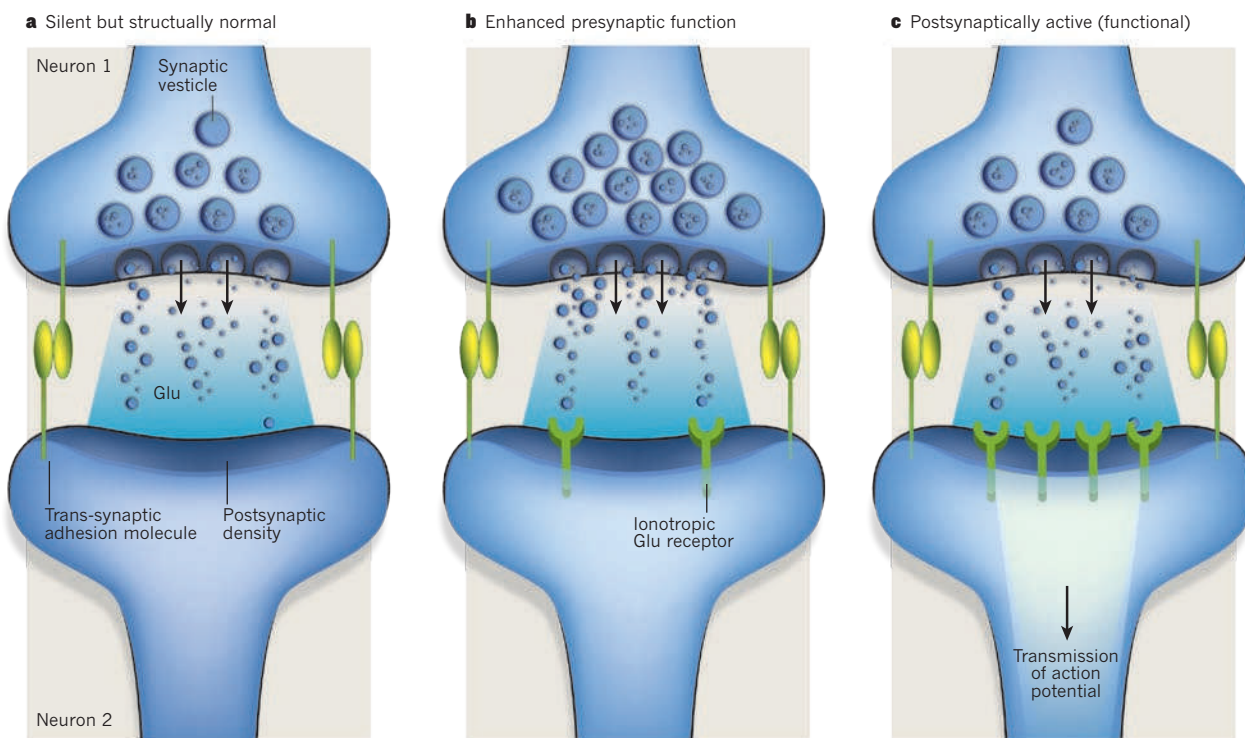
Figure 2 | Glial regulation of synaptic development. Several studies using the retinal ganglion cell culture system have shown that there are at least three classes of factor secreted by astrocytes. These factors control different aspects of the development of glutamate-mediated synapses. **a**, One type induces the formation of structurally normal but postsynaptically silent synapses. Thrombospondins are an example of this type of factor. **b**, Another type facilitates presynaptic activity and increases the probability of neurotransmitter release. Cholesterol functions in this way. **c**, A third type induces the formation of functional synapses or converts silent synapses into active ones by facilitating the insertion of glutamate receptors into postsynaptic sites. These factors have yet to be identified.

Factors secreted by astrocytes also regulate the formation of inhibitory synapses between hippocampal neurons *in vitro*. They do this by modulating inhibitory postsynaptic development, by stimulating neuron–neuron signalling through the neurotrophic receptor tyrosine kinase TrkB[49]. The exact mechanism is not clear; however, the neurons themselves are the source of the TrkB ligand brain-derived neurotrophic factor (BDNF). TSP1 did not induce inhibitory synapse formation in this culture system, although it facilitated excitatory synapse formation and neurite outgrowth[50]. The identity of the factor(s) that is secreted by astrocytes and promotes inhibitory synaptogenesis through increasing neuronal BDNF production remains unknown.

Schwann cells of the peripheral nervous system, similar to astrocytes in the CNS, also secrete synaptogenic factors. These factors include transforming growth factor-β, which controls neuromuscular junction formation[51], and unidentified low-molecular-mass molecules that increase synapse function[52]. In addition, both TSP4 and $\alpha_2\delta$-1 are highly enriched at the neuromuscular junction[53,54], suggesting that they might be important in inducing synapse formation in the peripheral nervous system, as well as in the CNS.

Astrocytes also regulate synapse formation through contact-mediated mechanisms. In the retina, there is a developmental switch in which the ability of RGCs to respond to soluble synaptogenic signals from astrocytes is induced by contact with astrocytes themselves[55]. This finding was made by observing rat RGCs at embryonic day (E) 17 co-cultured with postnatal RGCs in the presence of astrocyte-conditioned medium. Synapses did not form on the dendrites of E17 RGCs, whereas E19 RGCs received synapses. This two-day period (E17–E19) coincides with the developmental window *in vivo* during which astrocytes migrate along the optic nerve into the retina and populate the inner retinal layer adjacent to the RGCs. Contact with astrocytes but not amacrine cells, which are interneurons in the retina, was sufficient for cultured E17 RGCs to become receptive to synapse formation. Although the exact mechanism is not yet clear, it is known that local contact with astrocytes through integrins activates protein kinase C signalling in individual dissociated neurons from the embryonic hippocampus, and facilitates excitatory synapse formation[56]. It is possible that this global protein kinase C activation initiates neuronal maturation events that affect major cellular processes in neurons, such as the sorting of synaptic adhesion proteins into the correct subcellular compartments. Consistent with this hypothesis, it has been shown that astrocyte contact with E17 RGCs stimulates the localization of the synaptic adhesion molecule neurexin away from dendrites[55], where neurexin functions as an inhibitor of synapse formation by interacting in *cis* with the postsynaptic adhesion molecule neuroligin[57].

Recent gene-expression analysis of astrocytes *in vivo* showed that astrocytes produce messenger RNAs that encode several synaptic adhesion proteins, such as neurexins, neuroligins and cadherins[58], which are known to be important for synapse formation and are thought to function only in neurons[34]. Astrocyte processes might also use these cell-surface molecules to guide synapse formation or mediate the astrocyte–synapse interactions discussed earlier. In agreement with this, γ-protocadherins, a family of neuronal adhesion molecules encoded by a single gene cluster of 22 genes, are also expressed by astrocytes. These adhesion molecules localize to perisynaptic astrocyte processes, and homophilic astrocyte–neuron γ-protocadherin contacts are crucial for synaptogenesis *in vitro*. *In vivo*, restricted mutation of the γ-protocadherin gene cluster in astrocytes significantly delays the formation of both excitatory and inhibitory synapses[59]. These results suggest that cell adhesion molecules that are important for neuron–neuron interactions can also participate in astrocyte–neuron interactions, thus enabling astrocytes to guide synapse formation and morphology.

There is evidence that the glia–neuron signalling observed to regulate synapse formation in mammals also occurs in invertebrates. For example, in *Caenorhabditis elegans*, the glia-derived cell-surface receptor UNC-6 (also known as netrin) is used not only for guiding axons but also for defining sites for synaptogenesis[60]. In addition, the function
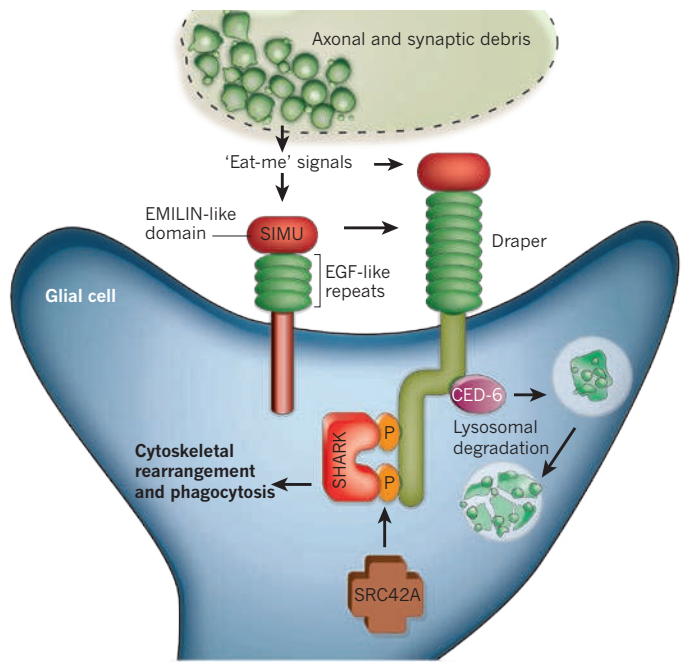
Figure 3 | Molecular pathways known to regulate axon pruning and synapse elimination by glia in invertebrates. Two transmembrane proteins, Six-microns-under (SIMU) and Draper, regulate the engulfment and phagocytosis of axosomes by glia in *Drosophila*. These proteins are homologues and have large extracellular regions with multiple epidermal growth factor (EGF)-like repeats and an EMILIN-like domain. SIMU and Draper function in the same pathway, most probably recognizing synaptic debris by binding to unidentified 'eat-me' signals on degenerating axons. SIMU seems to be involved in the initial recognition and uptake steps of engulfment, but it lacks an intracellular signalling domain. By contrast, Draper is capable of intracellular signalling and operates downstream of SIMU. The cytoplasmic adaptor protein CED-6 functions downstream of Draper, mediating the internalization and lysosomal degradation of debris. In addition, Draper triggers cytoskeletal rearrangements and phagocytosis through an interaction with the non-receptor tyrosine kinase SHARK. The kinase SRC42A facilitates the SHARK–Draper interaction by phosphorylating Draper.

of sensory neurons in *C. elegans* depends on FIG-1, a protein that is secreted by glia and contains TSP type I and type II EGF-like repeats[61]. Moreover, a *Drosophila* homologue of $\alpha_2\delta$-1, Cacna2d3, is required for presynaptic maturation at the neuromuscular junction[62]. Together, these findings suggest that the molecular basis of glia-induced synapse formation, as well as synapse formation itself, might be highly evolutionarily conserved. Interestingly, two isoforms of TSP (TSP2 and TSP4) are among the few genes that are substantially upregulated in the human brain compared with the primate and mouse brain[63], suggesting that TSPs — and thus astrocytes — might contribute to the greater brain plasticity of humans.

Several recent studies have shown that glia, through their ability to signal to neurons, also have an active role in activity-dependent modulation of synaptic efficacy. This process contributes to neural circuit development and experience-dependent plasticity. Regulation of synapse plasticity by long-term potentiation and long-term depression requires rapid adjustments in the strength of individual synapses in response to patterns of correlated synapse activity. The main mechanisms regulating these processes involved in plasticity are thought to be changes in the cell-surface delivery, retention and Ca^{2+}-channel properties of the postsynaptic ionotropic glutamate receptors, AMPA receptors and NMDA (*N*-methyl-D-aspartate) receptors[64,65]. NMDA receptors contain a glutamate-binding site, as well as a site for binding to glycine, a co-agonist that is required for the opening of the Ca^{2+} channel. The amino acid D-serine, which is produced exclusively by astrocytes, is a potent agonist for this site[66]. In hippocampal cultures, neuronal-activity-
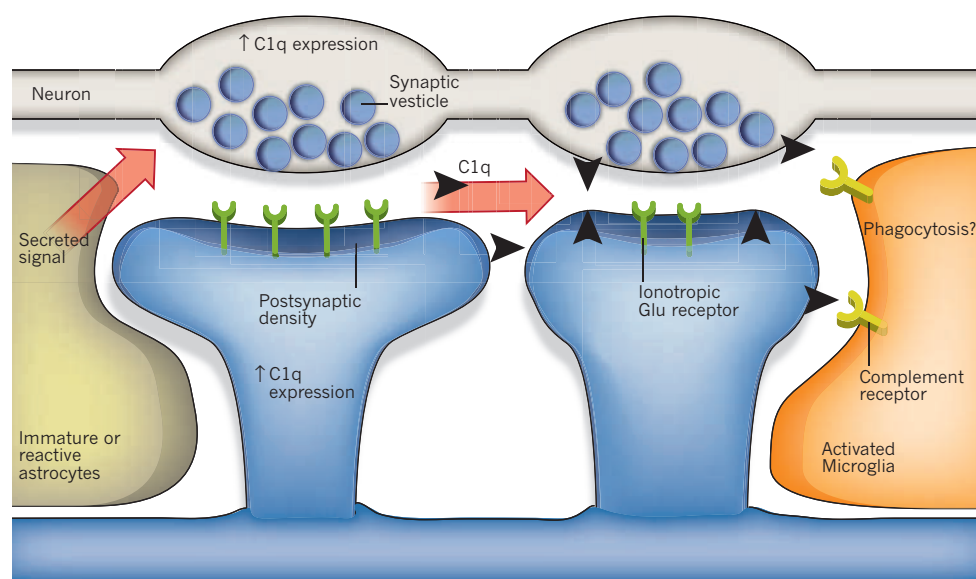
Figure 4 | Regulation of synapse elimination in the mammalian CNS by the complement cascade. An unidentified secreted signal from immature and reactive astrocytes upregulates expression of the complement component C1q in neurons. It is proposed that C1q binds to weaker synapses and tags them for elimination. This elimination might occur through phagocytosis by microglia, mediated by complement receptors at the surface of microglia. Other complement-cascade components such as C3 are also produced by glia in normal and disease conditions, and another possibility is that synapse elimination is triggered by C1q and C3 not only during development but also during the early stages of neurodegenerative diseases.

dependent release of D-serine from astrocytes was found to be necessary for long-term potentiation[67]. Moreover, in the hypothalamic supraoptic nucleus, a system in which structural changes in astrocyte association with synapses are important for oxytocin-mediated responses at synapses, the endogenous co-agonist of NMDA receptors is D-serine rather than glycine. The degree of astrocyte coverage of neurons governs the level at which D-serine occupies the glycine-binding site on the NMDA receptor, thereby controlling the activity dependence of long-term potentiation and long-term depression[68].

In addition to NMDA receptor function, cell-surface delivery of AMPA receptors is also regulated by the pro-inflammatory cytokine tumour-necrosis factor-α (TNF-α), which is released by glia. Glial TNF-α has been shown to improve synaptic efficacy by increasing the cell-surface expression of AMPA receptors. Conversely, blocking TNF-α-mediated signalling was found to have the opposite effect[69]. Interestingly, even though TNF-α increased the cell-surface expression of AMPA receptors, long-term potentiation and long-term depression were not affected in its absence[70]. Prolonged changes in the cell's synaptic activity, such as during blockade of synapse function, lead to adjustments in the strength of all synapses on that cell. This form of synaptic plasticity is called homeostatic synaptic scaling and is thought to be a crucial mechanism for preventing neural networks from becoming unstable and dysfunctional[71]. Glial TNF-α mediates synaptic scaling in response to prolonged activity blockade[70]. The exact glial-cell source of TNF-α (astrocytes or microglia) is not yet clear. Because TNF-α is secreted by microglia in response to injury or insult to the CNS, TNF-α-mediated synaptic plasticity might modulate neural responses to injury and neurodegeneration.

In summary, secreted and cell-surface-associated signals from glia coordinate excitatory and inhibitory synapse formation and modulate synapse function and plasticity in the CNS and peripheral nervous system (Table 1).

Glia regulate axon pruning and synapse elimination

During development, neurons often extend their axons beyond their intended targets and form an excessive number of synapses. The selective elimination of synapses and the pruning of axons to fine-tune synaptic territories are crucial for the proper development and function of the nervous system. The specific molecular mechanisms that drive synapse elimination and axon pruning remain mostly unknown, although, in the past decade, studies in invertebrate systems have suggested that synapses are removed actively through engulfment and phagocytosis by glia. Recent evidence suggests that vertebrate glia are also actively involved in the process of synapse elimination through mechanisms that

are both similar and distinct from those observed in invertebrates.

Developmental axon pruning is widely used in *Drosophila* metamorphosis, which involves major pruning events that require the breakdown and reconstruction of synaptic networks[72]. For example, in *Drosophila* larvae, γ-neurons in the mushroom body initially extend axon branches into both the dorsal and the medial mushroom-body-axon lobes. Axon branches to both lobes degenerate during pupal stages before the formation of adult connections. This pruning is triggered in a developmentally regulated manner by intrinsic molecular mechanisms such as activation of local axon degeneration by the neuronal nuclear hormone ecdysone[73]. Live imaging studies have shown that during early pupal stages, the processes of the glia that surround the neurons infiltrate bundles of axon branches. Glial processes engulf clusters of axon varicosities, which accumulate in intracellular lysosomal compartments. Selective inhibition of endocytosis by glia or of the ecdysone receptors in mushroom-body neurons suppresses infiltration by glia, as well as elimination of varicosities, and induces a severe delay in axon pruning[73,74]. These findings show that glia are activated by mushroom-body axons that are destined for elimination. Activated glia then infiltrate the mass of axon branches to eliminate varicosities and break down axon branches, rather than just scavenging already-degraded axon debris.

Forward genetic screens in *Drosophila* uncovered two genes — encoding Draper and CED-6, which are essential for the clearance of apoptotic cells in *C. elegans* — that function in the engulfment of pruned axons by glia during *Drosophila* metamorphosis (Fig. 3). In flies in which the gene encoding *draper* is mutated, or in which *draper* and *ced-6* have been knocked down specifically in glia by RNA interference, engulfment by glia is suppressed, resulting in the inhibition of axon pruning during metamorphosis[75,76]. These findings suggest that a similar molecular mechanism governs the clearance of apoptotic cells and of degenerating axons of living neurons. But how does Draper recognize modified self proteins? Another *Drosophila* receptor involved in phagocytosis, Six-microns-under (SIMU; also known as NIMC4), was recently reported to be required for efficient clearance of apoptotic cells by glia in the nervous system and by macrophages elsewhere[77]. SIMU is highly expressed by phagocytic cell types during development and is part of a conserved family of proteins that includes Draper. SIMU functions upstream of Draper in the same pathway and affects the recognition and engulfment of apoptotic cells, but only Draper affects their subsequent degradation[77] (Fig. 3). These results suggested that, by strongly binding to apoptotic cells or axons that are to be pruned, SIMU couples debris from degenerating axons to Draper signalling, which induces glia to phagocytose the debris.

Signalling of neural injury in *Drosophila* uses the same molecular

and cellular processes observed in developmental axon pruning. Upon axonal injury, glia upregulate the expression of *draper*, undergo marked changes in morphology and rapidly extend fine processes towards severed axons[78]. In *draper* mutants, glia fail to respond to axon injury, and severed axons are not cleared from the CNS. The *Drosophila* protein SHARK, a non-receptor tyrosine kinase similar to mammalian SYK, has been identified as a downstream signalling molecule that binds to the intracellular domain of Draper (Fig. 3). SHARK activity is essential for Draper signalling events, including the recruitment of glial membranes to severed axons and the phagocytosis of axon debris and neuronal-cell corpses by glia. Another signalling molecule, the SRC-family kinase SRC42A, functions in the same pathway by increasing Draper phosphorylation, thus stimulating the phagocytic activity of glia[78] (Fig. 3). These Draper–SRC42A–SHARK interactions are also likely to govern developmental axon pruning and have remarkable similarities to immunoreceptor–SRC-family-kinase–SYK signalling in mammalian immune cells.

Glia also play important roles in synapse elimination in the mammalian nervous system. One of the classic examples of activity-dependent synapse elimination occurs at the mammalian neuromuscular junction. At birth, postsynaptic muscle cells are innervated by multiple motor axons. By the second week after birth, activity-dependent competition permanently eliminates immature inputs, whereas the sole remaining input is maintained and strengthened. Eliminated connections detach from the neuromuscular junction and, as they retract, pieces of axon are shed. In the mammalian peripheral nervous system, similar to the active engulfment and clearance by glia that is observed in *Drosophila* metamorphosis, Schwann cells break up retracting axons and remove the synaptic debris. Time-lapse imaging and serial electron microscopy demonstrated that as inappropriate axons disappeared, they shed small membrane particles that contained intact presynaptic structures[79]. These debris structures, termed axosomes, were formed by engulfment of the tips of retreating axons by neighbouring Schwann cells. Further time-lapse imaging was carried out in transgenic mice that differentially express fluorescent proteins in Schwann cells and axons. This experiment directly showed that axosome shedding occurs entirely within the confines of Schwann cells and depends on glial lysosome function[80]. The molecules that drive Schwann cells to phagocytose retracting axons are unknown. However, a recent study showed that, similar to mammals, during neuromuscular junction development in *Drosophila*, there is activity-dependent elimination of immature synaptic boutons and widespread appearance of presynaptic debris[81]. Moreover, like Schwann cells, *Drosophila* glia invade the neuromuscular junction and, together with

muscle cells, phagocytose the synaptic debris. The Draper signalling pathway is important for this process because suppression of Draper function results in accumulation of presynaptic debris and compromised synapse growth[81]. These results show that the activity-dependent elimination of presynaptic inputs in the neuromuscular junction through phagocytosis by glia is an evolutionarily conserved phenomenon. The Draper–CED-6 pathway might also be important as a regulator of glia-mediated axon pruning at the mammalian neuromuscular junction because peripheral glia express a similar or identical pathway[82].

A recent study showed that the classical complement cascade, which is part of the innate immune system, helps to mediate synapse elimination in the developing CNS[83]. During development, immature astrocytes produce an unidentified signal that alters gene expression in neurons, leading to the upregulated expression of C1q, the protein that initiates the complement cascade. Early in development, C1q is localized to synapses that are destined for elimination (Fig. 4). In support of this model, mice deficient in either C1q or the downstream complement protein C3 fail to execute synapse elimination properly in the CNS, as determined by the absence of anatomical refinement of retinogeniculate connections and by the retention of excess retinal innervations by lateral geniculate neurons. Moreover, this failure to eliminate synapses in C1q-deficient mice led to enhanced synaptic connectivity and epileptic activity in the adult mice[84], suggesting that defects in this pathway could be involved in seizure disorders observed in humans.

These observations beg the question, how does C1q control synapse elimination? Looking to the innate immune system, C1q functions as a molecular tag to mark unwanted cells or debris for removal[85]. So it is possible that C1q tags weak synapses for elimination (Fig. 4). The molecular interactions that localize C1q specifically to weak synapses are unknown; however, microglia, the resident immune cells of the brain, produce large amounts of the receptors for C1q and C3 and thus are likely to be responsible for the removal of unwanted synapses. Microglia also phagocytose the synaptic terminals of motor neurons after injury, through a process known as synaptic stripping (discussed in the next section)[86], and an exciting question is whether synaptic stripping is also complement dependent. Interestingly, homologues of C1q, such as precerebellins[87] and C1q-like protein 1 (C1QL1), C1QL2 and C1QL3 (ref. 88), are synaptic molecules that are expressed by neurons and are involved in activity-dependent synapse plasticity and formation. It is thus likely that these molecules all participate in a common mechanism that controls different aspects of synapse remodelling. Perhaps these synaptic C1q-like molecules compete for a common synaptic receptor[89], and their interaction with this receptor might be antagonized

Table 1 | Glial-cell signals that control synapse development

Glia type	Signal type	Signal	Action
Rat astrocyte	Secreted	TSP1, TSP2, TSP3, TSP4 and TSP5	Induce excitatory synapse formation (but synapses are postsynaptically silent, owing to a lack of AMPA receptors)[41,42]
		Unknown	Induces excitatory synapse formation (and synapses are postsynaptically active, given the presence of AMPA receptors at the synapse)
		Unknown	Induces inhibitory synapse formation and neurite outgrowth by activating TrkB signalling in neurons[49,50]
		Cholesterol and apolipoprotein E	Enhance synaptic function[47] by facilitating presynaptic release[41] and dendritic maturation[48]
		D-Serine	Controls activity dependence of long-term potentiation and long-term depression[68]
	Cell surface	Unknown, possibility integrins (PKC-signalling dependent)	Regulates synaptic receptivity of embryonic neurons[55] in response to contact-dependent astrocyte signals[56]
		γ-Protocadherins	Facilitate correct execution of excitatory and inhibitory synaptogenesis[59]
Frog Schwann cell	Secreted	TGF-β	Induces neuromuscular junction formation[51]
		Unidentified low-molecular-mass molecule (not ATP or glutamate)	Enhances synaptic activity at the neuromuscular junction[52]
C. elegans glia	Secreted	FIG-1	Regulates sensory organ formation and function[61]
	Cell surface	UNC-6 (also known as netrin)	Instructs axonal growth and presynaptic assembly[60]

by C1q. The recent finding that the synaptic pentraxins, which bind to C1q, mediate silent to active synapse conversion in the developing visual system[90], potentially provides an exciting link between activity-dependent control of synapse elimination and regulation of synapse function, given that neuronal pentraxins are implicated in postsynaptic glutamate-receptor recruitment.

Glia control synaptic connectivity in disease

Glia rapidly respond to injury in the nervous system. The activation of microglia is triggered by injury-mediated signals, such as ATP released from dead cells and serum factors leaking into the extracellular environment as a result of the breakdown of the blood–brain barrier. Similar signalling events also trigger a marked change in astrocyte behaviour. Both cell types lose their ramified structures. Microglia migrate to the injury site and start to divide, and astrocytes direct their processes towards the damaged region[86,91]. This phenomenon is known as reactive gliosis. The mechanisms that affect the behaviour of glia after insult to the CNS are an active area of research. Whether reactive glia aid in CNS recovery after injury or permanently impair CNS regeneration is not clear.

When injured, reactive astrocytes revert to an immature state and express molecules that affect synapse formation. For example, the production of TSP1 and TSP2 increases after traumatic brain injury, ischaemia or stroke[92]. Purinoceptor signalling and mechanical stimulation have been shown to mediate the upregulation of TSP production in cultured astrocytes, and these proteins could be upregulated in a similar manner *in vivo* after injury[93]. TSP1 and TSP2 have been shown to be required for functional recovery after stroke because mice deficient in both TSP1 and TSP2 (TSP1/2-nulls) showed impaired recovery of motor function, synaptic density and axon sprouting[94]. Another line of evidence suggests that the TSP–$\alpha_2\delta$-1 interaction, and thus astrocyte-induced synapse formation, is necessary for the correct execution of injury-mediated developmental plasticity in the mouse brain[42]. This was determined by using a well-characterized model of barrel-cortex plasticity. Inhibition of TSP-mediated synapse formation, either by injection of gabapentin or in TSP1/2-null mice, strongly perturbed the stereotypical reorganization of the barrel cortex.

TSP–$\alpha_2\delta$-1 signalling after injury might not always be beneficial. The fact that the anti-analgesic, anti-epileptic drug gabapentin blocks the interaction between $\alpha_2\delta$-1 and TSP suggests that under certain conditions this signalling might be pathological. In agreement with this idea, both Cacna2d1 and TSP4 levels increase in the spinal-nerve ligation model of neuropathic pain[95,96]. The increase in the expression levels of Cacna2d-1 and TSPs could be an essential part of synapse remodelling in response to synapse loss after injury. However, aberrant synapse formation — or dysregulation of another cellular process that is also regulated by the TSP–$\alpha_2\delta$-1 interaction (for example, increased trafficking of Ca^{2+} channels to the cell surface) — may lead to outcomes such as epilepsy and neuropathic pain. If this is the case, then drugs that efficiently target the TSP–$\alpha_2\delta$-1 interaction may help to alleviate neuropathic pain or even to prevent the development of epilepsy following traumatic brain injury.

Similar to the engulfing glia of *Drosophila* that were discussed earlier, microglia and possibly astrocytes physically remove synaptic inputs after injury. For example, after axotomy, activated microglia adhere to damaged motor neurons, spread across their soma and dendrites, and actively denervate (remove neuron–neuron (nerve) connections in a way that breaks down damaged synapses) the glutamate-containing presynaptic boutons[86]. This process, generally referred to as synaptic stripping, is thought to protect neurons from excitotoxicity and increase the overall inhibition of the damaged neuron. There are strong indications that signalling between neurons and glia through receptors that are classically involved in immune responses underlies this stripping response. Inhibition of classical MHC class I molecule signalling led to an aberrant stripping response by microglia such that not only excitatory synapses were removed in response to injury but also a subclass of inhibitory synapses[97].

Synaptic stripping could also be harmful. Emerging evidence suggests that the loss of synapses is an early hallmark of neurodegenerative diseases, including Alzheimer's disease, glaucoma and prion disease[85,86,91]. Interestingly, components of the complement system, such as C1q, the C1 complex and activated C3, have been shown to localize to senile plaques in patients with Alzheimer's disease but were not present in the brains of a control group of non-demented, elderly individuals. Moreover, activated microglia and astrocytes are present in the vicinity of these plaques, indicating that excessive synaptic stripping might be occurring at these locations. Early in the course of a disease process, it might be useful to strip away dying synapses so that healthy synapses can take over this synaptic territory. It has been postulated that C1q and C3, which may initially be protective, are differentially synthesized early in neurodegenerative diseases. However, as the insult to the CNS persists, the remainder of the components of the complement cascade are synthesized, resulting in synapse loss and cell death[85] (Fig. 4). In agreement with this idea, C1q has been shown to be synthesized in response to injury or during the early stages of neurodegeneration in the adult CNS. For example, in the mouse model of glaucoma, C1q production was found to be upregulated, and the protein was localized to RGC synapses in the inner plexiform layer of the retina; however, this occurred only during the early stages of the disease, preceding the substantial synapse loss and eventual RGC loss that is observed in this model[83]. These findings suggest that C1q-dependent synapse loss may be driving the neurodegenerative process in glaucoma and other neurodegenerative disorders. This exciting finding could be further investigated by determining whether C1q deficiency is protective against neurodegenerative disease in animal models.

Future directions and outstanding questions

The exciting new findings highlighted here have vastly advanced the understanding of the cellular and molecular mechanisms that glia use to achieve correct synaptic connectivity in the nervous system. But many questions and mysteries remain unsolved. For example, even though the importance and diversity of astrocytes is now recognized, the full extent of their properties and the specific roles they have in the formation and function of the CNS are still unclear. Why are astrocytes tiled into non-overlapping anatomical domains? What are the functional consequences of localized and global changes in intracellular Ca^{2+} concentrations in these cells in response to synapse function? Could astrocytes and other glia actively influence neural circuits to control information processing? Do microglia have an active role in synapse elimination during development? Furthermore, the precise function of OPCs in the adult brain still needs to be uncovered, as well as the purpose and consequences of their direct synaptic contacts with neurons. The good news is that many of the technical difficulties that impeded glial research in the past are, to a great extent, solved. Recent advances in methods to carry out cell-type-specific genetic manipulations in invertebrate and vertebrate model organisms, as well as powerful *in vivo* imaging techniques, now allow the study and observation of glial function at synapses in great detail. Today, neuroscientists are in a better position than ever before for exploring and uncovering the long-standing mysteries of glia and for gaining new insight into the general workings of the nervous system. ∎

1. Barres, B. A. The mystery and magic of glia: a perspective on their roles in health and disease. *Neuron* **60**, 430–440 (2008).
2. Chittajallu, R., Aguirre, A. & Gallo, V. NG2-positive cells in the mouse white and grey matter display distinct physiological properties. *J. Physiol. (Lond.)* **561**, 109–122 (2004).
3. Lin, S. C. & Bergles, D. E. Physiological characteristics of NG2-expressing glial cells. *J. Neurocytol.* **31**, 537–549 (2002).
4. Eroglu, C., Barres, B. A. & Stevens, B. in *Structural and Functional Organization of the Synapse* (eds Hell, J. W. & Ehlers, M. D.) 683–714 (Springer, 2008).
5. Feng, Z. & Ko, C. P. Neuronal glia interactions at the vertebrate neuromuscular junction. *Curr. Opin. Pharmacol.* **7**, 316–324 (2007).
6. Bolton, M. M. & Eroglu, C. Look who is weaving the neural web: glial control of synapse formation. *Curr. Opin. Neurobiol.* **19**, 491–497 (2009).
7. Reichenbach, A., Derouiche, A. & Kirchhoff, F. Morphology and dynamics of

perisynaptic glia. *Brain Res. Rev.* **63**, 11–25 (2010).

8. Doherty, J., Logan, M. A., Tasdemir, O. E. & Freeman, M. R. Ensheathing glia function as phagocytes in the adult *Drosophila* brain. *J. Neurosci.* **29**, 4768–4781 (2009).

9. Dani, J. W., Chernjavsky, A. & Smith, S. J. Neuronal activity triggers calcium waves in hippocampal astrocyte networks. *Neuron* **8**, 429–440 (1992).

10. Wang, X. *et al.* Astrocytic Ca²⁺ signaling evoked by sensory stimulation *in vivo*. *Nature Neurosci.* **9**, 816–823 (2006).
    This paper shows that the stimulation of whiskers increases the cytosolic Ca²⁺ concentration in astrocytes in the barrel cortex of adult mice. This is the first *in vivo* evidence that astrocytes respond to a presynaptic spillover of glutamate by increasing their cytosolic Ca²⁺ concentrations.

11. Araque, A., Parpura, V., Sanzgiri, R. P. & Haydon, P. G. Tripartite synapses: glia, the unacknowledged partner. *Trends Neurosci.* **22**, 208–215 (1999).

12. Bergles, D. E., Roberts, J. D., Somogyi, P. & Jahr, C. E. Glutamatergic synapses on oligodendrocyte precursor cells in the hippocampus. *Nature* **405**, 187–191 (2000).
    This is the first report of glutamate-mediated synaptic contacts and fast synaptic transmission between pyramidal neurons and OPCs in the mammalian hippocampus.

13. Lin, S. C. *et al.* Climbing fiber innervation of NG2-expressing glia in the mammalian cerebellum. *Neuron* **46**, 773–785 (2005).

14. De Biase, L. M., Nishiyama, A. & Bergles, D. E. Excitability and synaptic communication within the oligodendrocyte lineage. *J. Neurosci.* **30**, 3600–3611 (2010).

15. Lin, S. C. & Bergles, D. E. Synaptic signaling between GABAergic interneurons and oligodendrocyte precursor cells in the hippocampus. *Nature Neurosci.* **7**, 24–32 (2004).

16. Zhang, Y. & Barres, B. A. Astrocyte heterogeneity: an underappreciated topic in neurobiology. *Curr. Opin. Neurobiol.* **20**, 588–594 (2010).

17. Oberheim, N. A. *et al.* Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276–3287 (2009).

18. Bushong, E. A., Martone, M. E., Jones, Y. Z. & Ellisman, M. H. Protoplasmic astrocytes in CA1 stratum radiatum occupy separate anatomical domains. *J. Neurosci.* **22**, 183–192 (2002).

19. Halassa, M. M., Fellin, T., Takano, H., Dong, J. H. & Haydon, P. G. Synaptic islands defined by the territory of a single astrocyte. *J. Neurosci.* **27**, 6473–6477 (2007).

20. Ventura, R. & Harris, K. M. Three-dimensional relationships between hippocampal synapses and astrocytes. *J. Neurosci.* **19**, 6897–6906 (1999).

21. Grosche, J. *et al.* Microdomains for neuron–glia interaction: parallel fiber signaling to Bergmann glial cells. *Nature Neurosci.* **2**, 139–143 (1999).

22. Witcher, M. R., Kirov, S. A. & Harris, K. M. Plasticity of perisynaptic astroglia during synaptogenesis in the mature rat hippocampus. *Glia* **55**, 13–23 (2007).

23. Genoud, C. *et al.* Plasticity of astrocytic coverage and glutamate transporter expression in adult mouse cortex. *PLoS Biol.* **4**, e343 (2006).

24. Hirrlinger, J., Hulsmann, S. & Kirchhoff, F. Astroglial processes show spontaneous motility at active synaptic terminals *in situ*. *Eur. J. Neurosci.* **20**, 2235–2239 (2004).

25. Theodosis, D. T. *et al.* Oxytocin and estrogen promote rapid formation of functional GABA synapses in the adult supraoptic nucleus. *Mol. Cell. Neurosci.* **31**, 785–794 (2006).

26. Iino, M. *et al.* Glia–synapse interaction through Ca²⁺-permeable AMPA receptors in Bergmann glia. *Science* **292**, 926–929 (2001).
    This paper shows that Ca²⁺-permeable AMPA receptors on glia are required for proper structural and functional relationships between Bergmann glia and glutamate-mediated synapses in the cerebellum.

27. Haber, M., Zhou, L. & Murai, K. K. Cooperative astrocyte and dendritic spine dynamics at hippocampal excitatory synapses. *J. Neurosci.* **26**, 8881–8891 (2006).

28. Nishida, H. & Okabe, S. Direct astrocytic contacts regulate local maturation of dendritic spines. *J. Neurosci.* **27**, 331–340 (2007).

29. Murai, K. K., Nguyen, L. N., Irie, F., Yamaguchi, Y. & Pasquale, E. B. Control of hippocampal dendritic spine morphology through ephrin-A3/EphA4 signaling. *Nature Neurosci.* **6**, 153–160 (2003).

30. Carmona, M. A., Murai, K. K., Wang, L., Roberts, A. J. & Pasquale, E. B. Glial ephrin-A3 regulates hippocampal dendritic spine morphology and glutamate transport. *Proc. Natl Acad. Sci. USA* **106**, 12524–12529 (2009).

31. Filosa, A. *et al.* Neuron–glia communication via EphA4/ephrin-A3 modulates LTP through glial glutamate transport. *Nature Neurosci.* **12**, 1285–1292 (2009).

32. Hatton, G. I. Function-related plasticity in hypothalamus. *Annu. Rev. Neurosci.* **20**, 375–397 (1997).

33. Piet, R., Poulain, D. A. & Oliet, S. H. Modulation of synaptic transmission by astrocytes in the rat supraoptic nucleus. *J. Physiol. (Paris)* **96**, 231–236 (2002).

34. Fox, M. A. & Umemori, H. Seeking long-term relationship: axon and target communicate to organize synaptic differentiation. *J. Neurochem.* **97**, 1215–1231 (2006).

35. Fields, R. D. Myelination: an overlooked mechanism of synaptic plasticity? *Neuroscientist* **11**, 528–531 (2005).

36. Muller, C. M. & Best, J. Ocular dominance plasticity in adult cat visual cortex after transplantation of cultured astrocytes. *Nature* **342**, 427–430 (1989).

37. Meyer-Franke, A., Kaplan, M. R., Pfrieger, F. W. & Barres, B. A. Characterization of the signaling interactions that promote the survival and growth of developing retinal ganglion cells in culture. *Neuron* **15**, 805–819 (1995).

38. Pfrieger, F. W. & Barres, B. A. Synaptic efficacy enhanced by glial cells *in vitro*. *Science* **277**, 1684–1687 (1997).

39. Ullian, E. M., Sapperstein, S. K., Christopherson, K. S. & Barres, B. A. Control of synapse number by glia. *Science* **291**, 657–661 (2001).
    References 38 and 39 show that synapse formation and function are powerfully affected by astrocytes in culture. This is the first report that these processes are not only controlled by neuronal mechanisms but can also be directed by glia.

40. Wu, H. *et al.* Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. *Proc. Natl Acad. Sci. USA* **104**, 13821–13826 (2007).

41. Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421–433 (2005).
    Following on from the work by Ullian and colleagues³⁹, this paper identifies the extracellular matrix proteins TSPs as the synaptogenic factors that are released by astrocytes.

42. Eroglu, C. *et al.* Gabapentin receptor α2δ-1 is a neuronal thrombospondin receptor responsible for excitatory CNS synaptogenesis. *Cell* **139**, 380–392 (2009).
    In this study, the receptor for TSPs that is involved in synapse formation was identified to be the Ca²⁺ channel subunit α₂δ-1, which is also the receptor for the commonly prescribed pain medication gabapentin.

43. Rohrbough, J., Grotewiel, M. S., Davis, R. L. & Broadie, K. Integrin-mediated regulation of synaptic morphology, transmission, and plasticity. *J. Neurosci.* **20**, 6868–6878 (2000).

44. Shi, Y. & Ethell, I. M. Integrins control dendritic spine plasticity in hippocampal neurons through NMDA receptor and Ca²⁺/calmodulin-dependent protein kinase II-mediated actin reorganization. *J. Neurosci.* **26**, 1813–1822 (2006).

45. Xu, J., Xiao, N. & Xia, J. Thrombospondin 1 accelerates synaptogenesis in hippocampal neurons through neuroligin 1. *Nature Neurosci.* **13**, 22–24 (2010).

46. Taylor, C. P. Mechanisms of analgesia by gabapentin and pregabalin: calcium channel α₂-δ [Caᵥα₂-δ] ligands. *Pain* **142**, 13–16 (2009).

47. Mauch, D. H. *et al.* CNS synaptogenesis promoted by glia-derived cholesterol. *Science* **294**, 1354–1357 (2001).

48. Goritz, C., Mauch, D. H. & Pfrieger, F. W. Multiple mechanisms mediate cholesterol-induced synaptogenesis in a CNS neuron. *Mol. Cell. Neurosci.* **29**, 190–201 (2005).

49. Elmariah, S. B., Hughes, E. G., Oh, E. J. & Balice-Gordon, R. J. Neurotrophin signaling among neurons and glia during formation of tripartite synapses. *Neuron Glia Biol.* **1**, 1–11 (2005).

50. Hughes, E. G., Elmariah, S. B. & Balice-Gordon, R. J. Astrocyte secreted proteins selectively increase hippocampal GABAergic axon length, branching, and synaptogenesis. *Mol. Cell. Neurosci.* **43**, 136–145 (2009).

51. Feng, Z. & Ko, C. P. Schwann cells promote synaptogenesis at the neuromuscular junction via transforming growth factor-β1. *J. Neurosci.* **28**, 9599–9609 (2008).

52. Cao, G. & Ko, C. P. Schwann cell-derived factors modulate synaptic activities at developing neuromuscular synapses. *J. Neurosci.* **27**, 6712–6722 (2007).

53. Arber, S. & Caroni, P. Thrombospondin-4, an extracellular matrix protein expressed in the developing and adult nervous system, promotes neurite outgrowth. *J. Cell Biol.* **131**, 1083–1094 (1995).

54. Arikkath, J. & Campbell, K. P. Auxiliary subunits: essential components of the voltage-gated calcium channel complex. *Curr. Opin. Neurobiol.* **13**, 298–307 (2003).

55. Barker, A. J., Koch, S. M., Reed, J., Barres, B. A. & Ullian, E. M. Developmental control of synaptic receptivity. *J. Neurosci.* **28**, 8150–8160 (2008).

56. Hama, H., Hara, C., Yamaguchi, K. & Miyawaki, A. PKC signaling mediates global enhancement of excitatory synaptogenesis in neurons triggered by local contact with astrocytes. *Neuron* **41**, 405–415 (2004).

57. Taniguchi, H. *et al.* Silencing of neuroligin function by postsynaptic neurexins. *J. Neurosci.* **27**, 2815–2824 (2007).

58. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).

59. Garrett, A. M. & Weiner, J. A. Control of CNS synapse development by γ-protocadherin-mediated astrocyte–neuron contact. *J. Neurosci.* **29**, 11723–11731 (2009).

60. Colon-Ramos, D. A., Margeta, M. A. & Shen, K. Glia promote local synaptogenesis through UNC-6 (netrin) signaling in *C. elegans*. *Science* **318**, 103–106 (2007).

61. Bacaj, T., Tevlin, M., Lu, Y. & Shaham, S. Glia are essential for sensory organ function in *C. elegans*. *Science* **322**, 744–747 (2008).

62. Kurshan, P. T., Oztan, A. & Schwarz, T. L. Presynaptic α₂δ-3 is required for synaptic morphogenesis independent of its Ca²⁺-channel functions. *Nature Neurosci.* **12**, 1415–1423 (2009).

63. Caceres, M., Suwyn, C., Maddox, M., Thomas, J. W. & Preuss, T. M. Increased cortical expression of two synaptogenic thrombospondins in human brain evolution. *Cereb. Cortex* **17**, 2312–2321 (2007).

64. Cummings, J. A., Mulkey, R. M., Nicoll, R. A. & Malenka, R. C. Ca²⁺ signaling requirements for long-term depression in the hippocampus. *Neuron* **16**, 825–833 (1996).

65. Selig, D. K., Hjelmstad, G. O., Herron, C., Nicoll, R. A. & Malenka, R. C. Independent mechanisms for long-term depression of AMPA and NMDA responses. *Neuron* **15**, 417–426 (1995).

66. Miller, R. F. D-Serine as a glial modulator of nerve cells. *Glia* **47**, 275–283 (2004).

67. Yang, Y. *et al.* Contribution of astrocytes to hippocampal long-term potentiation through release of D-serine. *Proc. Natl Acad. Sci. USA* **100**, 15194–15199 (2003).

68. Panatier, A. *et al.* Glia-derived D-serine controls NMDA receptor activity and synaptic memory. *Cell* **125,** 775–784 (2006).

69. Beattie, E. C. *et al.* Control of synaptic strength by glial TNFα. *Science* **295,** 2282–2285 (2002).

70. Stellwagen, D. & Malenka, R. C. Synaptic scaling mediated by glial TNF-α. *Nature* **440,** 1054–1059 (2006).

71. Buckby, L. E., Jensen, T. P., Smith, P. J. & Empson, R. M. Network stability through homeostatic scaling of excitatory and inhibitory synapses following inactivity in CA3 of rat organotypic hippocampal slice cultures. *Mol. Cell. Neurosci.* **31,** 805–816 (2006).

72. Freeman, M. R. Sculpting the nervous system: glial control of neuronal development. *Curr. Opin. Neurobiol.* **16,** 119–125 (2006).

73. Awasaki, T. & Ito, K. Engulfing action of glial cells is required for programmed axon pruning during *Drosophila* metamorphosis. *Curr. Biol.* **14,** 668–677 (2004).

74. Watts, R. J., Schuldiner, O., Perrino, J., Larsen, C. & Luo, L. Glia engulf degenerating axons during developmental axon pruning. *Curr. Biol.* **14,** 678–684 (2004).
References 73 and 74 are the first reports that the pruning of axons during *Drosophila* metamorphosis is mediated by their active engulfment by glia.

75. MacDonald, J. M. *et al.* The *Drosophila* cell corpse engulfment receptor Draper mediates glial clearance of severed axons. *Neuron* **50,** 869–881 (2006).

76. Awasaki, T. *et al.* Essential role of the apoptotic cell engulfment genes *draper* and *ced-6* in programmed axon pruning during *Drosophila* metamorphosis. *Neuron* **50,** 855–867 (2006).

77. Kurant, E., Axelrod, S., Leaman, D. & Gaul, U. Six-microns-under acts upstream of Draper in the glial phagocytosis of apoptotic neurons. *Cell* **133,** 498–509 (2008).

78. Ziegenfuss, J. S. *et al.* Draper-dependent glial phagocytic activity is mediated by Src and Syk family kinase signalling. *Nature* **453,** 935–939 (2008).

79. Bishop, D. L., Misgeld, T., Walsh, M. K., Gan, W. B. & Lichtman, J. W. Axon branch removal at developing synapses by axosome shedding. *Neuron* **44,** 651–661 (2004).

80. Song, J. W. *et al.* Lysosomal activity associated with developmental axon pruning. *J. Neurosci.* **28,** 8993–9001 (2008).

81. Fuentes-Medel, Y. *et al.* Glia and muscle sculpt neuromuscular arbors by engulfing destabilized synaptic boutons and shed presynaptic debris. *PLoS Biol.* **7,** e1000184 (2009).

82. Wu, H. H. *et al.* Glial precursors clear sensory neuron corpses during development via Jedi-1, an engulfment receptor. *Nature Neurosci.* **12,** 1534–1541 (2009).

83. Stevens, B. *et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131,** 1164–1178 (2007).
This study describes an unexpected role for the complement cascade in the elimination of synapses during development and provides evidence that the expression of complement proteins is upregulated and their synaptic localization increases during the early stages of glaucoma.

84. Chu, Y. *et al.* Enhanced synaptic connectivity and epilepsy in C1q knockout mice. *Proc. Natl Acad. Sci. USA* **107,** 7975–7980 (2010).

85. Alexander, J. J., Anderson, A. J., Barnum, S. R., Stevens, B. & Tenner, A. J. The complement cascade: Yin–Yang in neuroinflammation — neuro-protection and -degeneration. *J. Neurochem.* **107,** 1169–1187 (2008).

86. Cullheim, S. & Thams, S. The microglial networks of the brain and their role in neuronal network plasticity after lesion. *Brain Res. Rev.* **55,** 89–96 (2007).

87. Urade, Y., Oberdick, J., Molinar-Rode, R. & Morgan, J. I. Precerebellin is a cerebellum-specific protein with similarity to the globular domain of complement C1q B chain. *Proc. Natl Acad. Sci. USA* **88,** 1069–1073 (1991).

88. Iijima, T., Miura, E., Watanabe, M. & Yuzaki, M. Distinct expression of C1q-like family mRNAs in mouse brain and biochemical characterization of their encoded proteins. *Eur. J. Neurosci.* **31,** 1606–1615 (2010).

89. Matsuda, K. *et al.* Cbln1 is a ligand for an orphan glutamate receptor δ2, a bidirectional synapse organizer. *Science* **328,** 363–368 (2010).

90. Koch, S. M. & Ullian, E. M. Neuronal pentraxins mediate silent synapse conversion in the developing visual system. *J. Neurosci.* **30,** 5404–5414 (2010).

91. Johnson, E. C. & Morrison, J. C. Friend or foe? Resolving the impact of glial responses in glaucoma. *J. Glaucoma* **18,** 341–353 (2009).

92. Lin, T. N. *et al.* Differential regulation of thrombospondin-1 and thrombospondin-2 after focal cerebral ischemia/reperfusion. *Stroke* **34,** 177–186 (2003).

93. Tran, M. D. & Neary, J. T. Purinergic signaling induces thrombospondin-1 expression in astrocytes. *Proc. Natl Acad. Sci. USA* **103,** 9321–9326 (2006).

94. Liauw, J. *et al.* Thrombospondins 1 and 2 are necessary for synaptic plasticity and functional recovery after stroke. *J. Cereb. Blood Flow Metab.* **28,** 1722–1732 (2008).

95. Li, C. Y., Song, Y. H., Higuera, E. S. & Luo, Z. D. Spinal dorsal horn calcium channel $\alpha_2\delta$-1 subunit upregulation contributes to peripheral nerve injury-induced tactile allodynia. *J. Neurosci.* **24,** 8494–8499 (2004).

96. Valder, C. R., Liu, J. J., Song, Y. H. & Luo, Z. D. Coupling gene chip analyses and rat genetic variances in identifying potential target genes that may contribute to neuropathic allodynia development. *J. Neurochem.* **87,** 560–573 (2003).

97. Oliveira, A. L. *et al.* A role for MHC class I molecules in synaptic plasticity and regeneration of neurons after axotomy. *Proc. Natl Acad. Sci. USA* **101,** 17843–17848 (2004).

REVIEW

Glial and neuronal control of brain blood flow

David Attwell[1], Alastair M. Buchan[2], Serge Charpak[3], Martin Lauritzen[4], Brian A. MacVicar[5] & Eric A. Newman[6]

Blood flow in the brain is regulated by neurons and astrocytes. Knowledge of how these cells control blood flow is crucial for understanding how neural computation is powered, for interpreting functional imaging scans of brains, and for developing treatments for neurological disorders. It is now recognized that neurotransmitter–mediated signalling has a key role in regulating cerebral blood flow, that much of this control is mediated by astrocytes, that oxygen modulates blood flow regulation, and that blood flow may be controlled by capillaries as well as by arterioles. These conceptual shifts in our understanding of cerebral blood flow control have important implications for the development of new therapeutic approaches.

The human brain comprises only 2% of the body's mass, but it consumes 20% of the energy that is produced when the body is in a resting state. This high consumption of energy is crucial for the normal functioning of the brain. The energy is mostly used to reverse the ion influxes that underlie synaptic potentials and action potentials[1] (Fig. 1a). If there is an inadequate supply of blood glucose and oxygen to a region of the brain, then neurons and glia become injured or die. This occurs in disorders such as ischaemic stroke, vasospasm after sub-arachnoid haemorrhage, the secondary ischaemia that follows spinal-cord injury, and cerebral palsy after perinatal asphyxia. To sustain neuronal function, the brain has evolved 'neurovascular coupling' mechanisms to increase the flow of blood to regions in which neurons are active, a response termed functional hyperaemia. Different information coding strategies and neural algorithms require different increases in blood flow, depending on the extent to which they consume energy. An understanding of the mechanisms that generate functional hyperaemia is a prerequisite for developing therapies to correct defects in blood flow control that occur after disorders such as stroke[2], hypertension[3], spinal-cord injury[4] and Alzheimer's disease[3].

Concepts of how neuronal activity controls the vascular supply of glucose and O_2 are changing rapidly. Traditionally, it was thought that active neurons generate a metabolic signal (a fall in O_2 or glucose concentration, or a rise in carbon dioxide concentration), which triggers an increase in blood flow. This idea has recently been superseded, following the discovery that neurotransmitter-mediated signalling, particularly by glutamate, has a major role in regulating cerebral blood flow, and that much of this control is mediated by astrocytes. Glutamate-mediated signalling leads to the release of nitric oxide from neurons and of arachidonic acid derivatives from astrocytes (and possibly from neurons). These molecules can either increase or decrease blood flow, depending on the local O_2 concentration, but how this switch occurs is debated. Furthermore, the relative importance of the different glutamate-released messengers varies between brain regions. Even the dogma that cerebral blood flow is controlled solely by arterioles has been challenged, with the finding that contractile cells called pericytes can control the diameter of capillaries, and that damage to these cells contributes to the long-lasting decrease in blood flow that occurs after stroke. These major conceptual shifts, which we discuss in

this Review, provide a new understanding of how the brain regulates its energy supply in response to different information processing tasks. They also underpin the interpretation of data from functional imaging experiments, and offer new opportunities for developing therapeutic approaches to a range of disorders of the central nervous system.

Neurotransmitters increase cerebral blood flow

In this Review, we focus on the control of cerebral blood flow by local neuronal activity. We pay less attention to autoregulation, which, in the face of changes in systemic blood pressure, maintains an approximately constant blood flow to the brain.

Early researchers favoured the idea that blood flow is locally controlled by a negative-feedback system in which neural activity leads to energy demand, because ATP is used to restore ion gradients after the generation of synaptic and action potentials[1]. This ATP use was thought to produce a metabolic signal that increased blood flow and therefore provided more energy (Fig. 1b). This metabolic signal could be a lack of O_2 or glucose, or the production of CO_2 (which dilates cerebral vessels by being converted to H^+ after combining with H_2O)[5].

However, manipulation of blood O_2 (refs 6, 7) and glucose[8] concentrations to test this negative-feedback hypothesis has shown that O_2 and glucose do not regulate blood flow in this way. Furthermore, during neuronal activity, the local extracellular pH initially becomes alkaline, rather than becoming acidic as would be expected if arteriolar dilation were caused by the accumulation of CO_2. This is partly because the increase in blood flow elicited by neural activity washes CO_2 away rapidly, and partly because the neuronal Ca^{2+}, H^+-ATPase pump alkalinizes the extracellular space when neural activity raises the neuronal intracellular calcium concentration ($[Ca^{2+}]_i$)[9,10]. These findings do not support a local CO_2 rise as the cause of the increased blood flow — despite the fact that exogenous CO_2 does increase blood flow. However, the 'metabolic messenger' adenosine (which is produced when ATP is hydrolysed) does contribute to functional hyperaemia, because blocking adenosine receptors reduces the increase in blood flow evoked by neuronal activity[11]. Furthermore, another metabolic messenger, lactate (which is produced when pyruvate production by glycolysis outstrips pyruvate consumption by oxidative phosphorylation), also increases

[1]Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1E 6BT, UK. [2]Acute Stroke Programme, University of Oxford, Nuffield Department of Medicine, Level 7, John Radcliffe Hospital, Oxford OX3 9DU, UK. [3]INSERM U603; CNRS UMR 8154; Laboratory of Neurophysiology and New Microscopies, Université Paris Descartes, Paris 75006, France. [4]Department of Clinical Neurophysiology, Institute for Neuroscience and Pharmacology, University of Copenhagen & Glostrup Hospital, Glostrup DK-2600, Denmark. [5]Brain Research Center & Department of Psychiatry, University of British Columbia, 2211 Westbrook Mall, Vancouver, British Columbia V6T 2B5, Canada. [6]Department of Neuroscience, University of Minnesota, 6-145 Jackson Hall, 321 Church Street SE, Minneapolis, Minnesota 55455, USA.
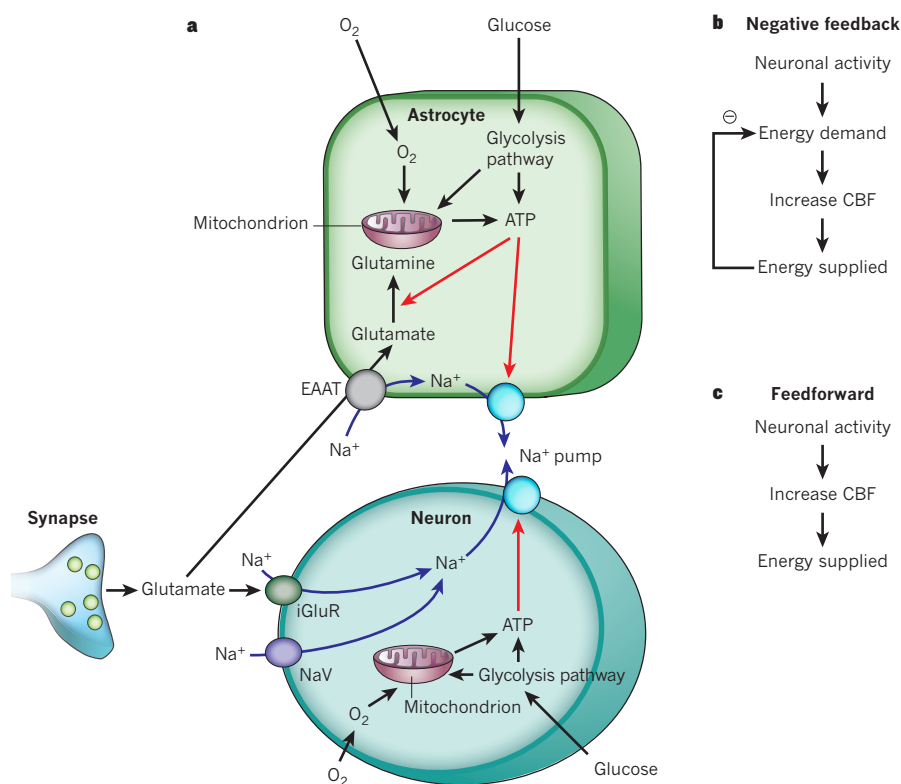
Figure 1 | Energy supply, usage and blood flow regulation in the brain. a, ATP is generated from glycolysis and mitochondrial oxidative phosphorylation in neurons and glia. ATP is mainly consumed (red arrows) by ion pumping in neurons, to maintain the ion gradients underlying synaptic and action potentials, following Na^+ entry (blue arrows) through ionotropic glutamate receptors (iGluR) and voltage-gated Na^+ channels (NaV). It is also used in glia for Na^+-coupled neurotransmitter uptake by excitatory amino acid transporters (EAAT) and for metabolic processing (shown for conversion of glutamate to glutamine), and on maintaining the cells' resting potentials. **b**, The negative-feedback control hypothesis for vascular energy supply, in which a fall in energy level induces an increased cerebral blood flow (CBF). **c**, The feedforward regulation hypothesis for vascular energy supply.

blood flow[12]. However, as explained below (see 'O$_2$ modulates neurovascular signalling'), this is mediated at least partly by a modulation of neurotransmitter-induced prostaglandin signalling to blood vessels.

More recent work has shown that control of the vascular energy supply by neural activity is largely mediated by feedforward mechanisms (Fig. 1c). In these processes, neurons either signal directly to blood vessels or activate astrocytes to release vasoactive agents onto the vessels. For both of these signalling routes, the coupling mechanisms involve neurotransmitter, particularly glutamate, signalling (Fig. 2).

One cannot distinguish between metabolic feedback and neurotransmitter feedforward mechanisms simply by blocking neurotransmitter-mediated signalling. Although neurotransmitter antagonists block the increase in blood flow elicited by neural activity[13–17], they also block the energy consumption evoked by synaptic signalling[1,18,19]. However, the roles of energy consumption and neurotransmitter-mediated signalling in regulating blood flow can be disentangled. Blocking the enzymes downstream of glutamate receptors that generate nitric oxide and arachidonic acid derivatives greatly reduces functional hyperaemia but has little effect on the energy use that is associated with neural activity[18,20].

Neuronal signalling to blood vessels

Synaptic release of glutamate activates neuronal NMDA (N-methyl-D-aspartate) receptors, resulting in Ca^{2+} entry into neurons and activation of neuronal nitric oxide synthase (nNOS). This releases NO, which dilates vessels[21], both in brain slices (which allow more refined mechanistic investigations, but lack intravascular pressure and blood flow that can release messengers that modulate the properties of vascular smooth muscle) and in vivo (Fig. 2). In the cortex, inhibition of nNOS reduces the increase in blood flow that is associated with neural activity[22], suggesting a role for NO in neurovascular coupling. However, the response is restored by the addition of NO donors, which provide a constant concentration of NO. This indicates that, although the presence of NO is required, a dynamic rise of NO concentration in response to neuronal activity does not directly mediate neuron-to-vessel signalling[23]. Instead, NO might be needed to modulate the pathways in astrocytes which dilate and constrict blood vessels (see 'NO modulates

astrocyte signalling'). In the cerebellum, by contrast, NO donors do not reverse the reduction in functional hyperaemia caused by inhibiting nNOS, demonstrating that NO directly mediates a component of the response[24,25]. The activation of peptidergic interneurons can also dilate or constrict vessels in the cortex[26], but the dilations that have been observed were usually irreversible (within the 7-min recording period). In addition, it is unclear whether these effects involve peptide release and whether they reflect direct signalling from neurons to vessels or indirect signalling by way of astrocytes (see 'Astrocyte signalling to blood vessels'). Finally, the neurotransmitter GABA (γ-aminobutyric acid), acting through GABA$_A$ receptors, also mediates a component of the vasodilation produced in the cortex by basal forebrain stimulation[27], but it is unclear whether this is a direct effect on the vasculature or is mediated by an effect on neurons or astrocytes.

Astrocyte signalling to blood vessels

Astrocytes are ideally situated to function as relay cells in neurovascular communication, as was suggested more than a century ago by Ramón y Cajal. They surround synapses and thus can be stimulated by neuronal activity, whereas their endfoot processes envelop blood vessels and can signal to the smooth muscle cells that control vessel diameter.

In theory, astrocytes can increase blood flow in response to neuronal activity by releasing potassium ions from their endfeet apposed to arterioles, because modest increases in extracellular K$^+$ concentration (up to ~10 mM [K$^+$]$_o$) hyperpolarize smooth muscle cells. This occurs because a raised [K$^+$]$_o$ increases the conductance of smooth muscle inward rectifier K$^+$ channels, and this effect outweighs the positive shift of the K$^+$ reversal potential, E_K, produced by the raised [K$^+$]$_o$ and leads to more outward current flowing (because the membrane potential is more positive than E_K). This hyperpolarization reduces the influx of Ca^{2+} through voltage-gated channels and dilates the vessels[28]. The 'K$^+$ siphoning' hypothesis holds that K$^+$ released from active neurons depolarizes astrocytes, leading to K$^+$ efflux from astrocyte endfeet[29]. This mechanism of neurovascular coupling has been tested by directly depolarizing glial cells while monitoring blood vessel diameter[30]. Depolarization fails to dilate vessels, demonstrating that K$^+$ siphoning does not contribute significantly to
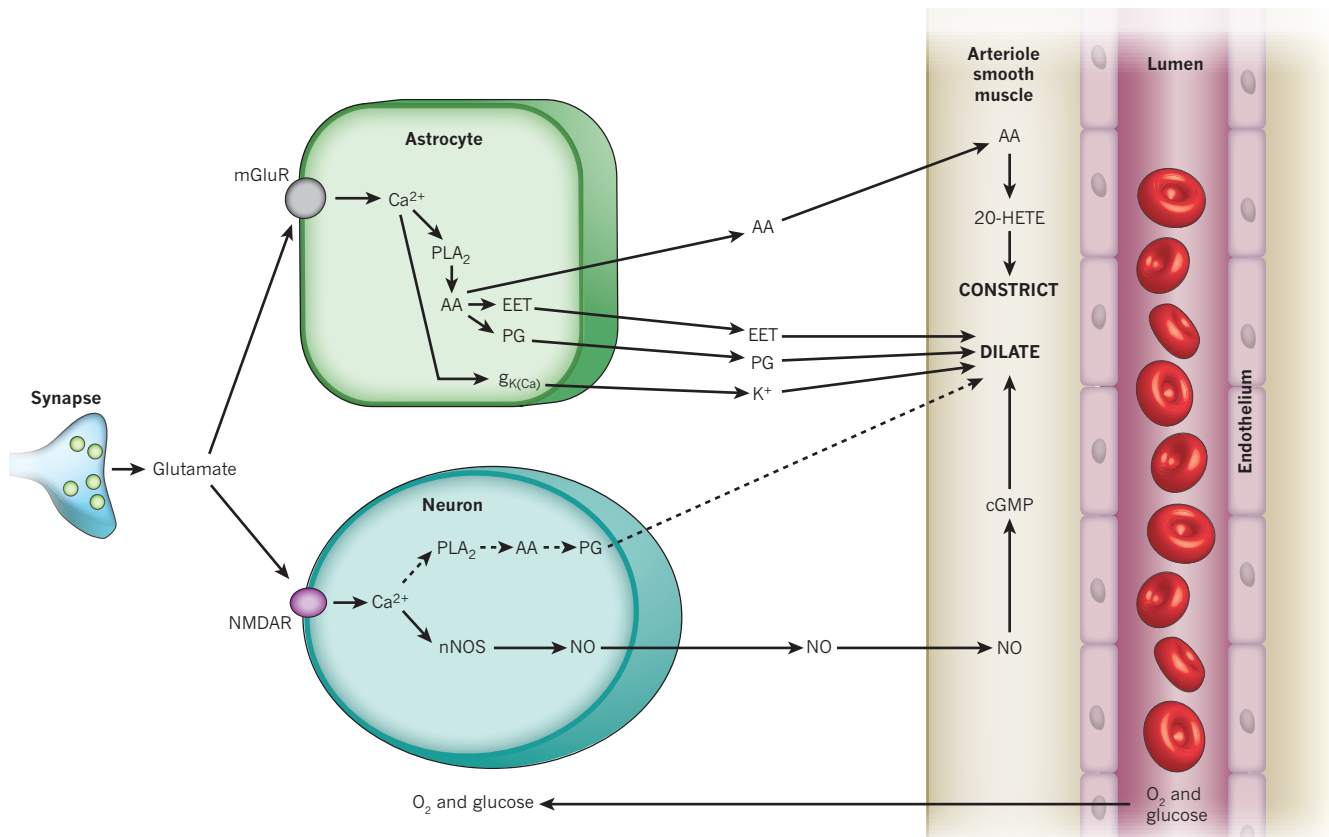
Figure 2 | Major pathways by which glutamate regulates cerebral blood flow. Pathways from astrocytes and neurons (left) that regulate blood flow by sending messengers (arrows) to influence the smooth muscle around the arterioles that supply oxygen and glucose to the cells (right, shown as the vessel lumen surrounded by endothelial cells and smooth muscle). In neurons, synaptically released glutamate acts on N-methyl-D-aspartate receptors (NMDAR) to raise $[Ca^{2+}]_i$, causing neuronal nitric oxide synthase (nNOS) to release NO, which activates smooth muscle guanylate cyclase. This generates cGMP to dilate vessels. Raised $[Ca^{2+}]_i$ may also (dashed line)

generate arachidonic acid (AA) from phospholipase A_2 (PLA_2), which is converted by COX2 to prostaglandins (PG) that dilate vessels. Glutamate raises $[Ca^{2+}]_i$ in astrocytes by activating metabotropic glutamate receptors (mGluR), generating arachidonic acid and thus three types of metabolite: prostaglandins (by COX1/3, and COX2 in pathological situations) and EETs (by P450 epoxygenase) in astrocytes, which dilate vessels, and 20-HETE (by ω-hydroxylase) in smooth muscle, which constricts vessels. A rise of $[Ca^{2+}]_i$ in astrocyte endfeet may activate Ca^{2+}-gated K^+ channels ($g_{K(Ca)}$), releasing K^+, which also dilates vessels.

vasodilation. However, astrocytes may dilate vessels through a different K^+-based mechanism. When neuronal activity releases glutamate at synapses, some of the released glutamate escapes the synaptic cleft and activates metabotropic glutamate receptors (mGluRs) on astrocytes, thus increasing $[Ca^{2+}]_i$ in astrocytes[31] (Fig. 2). This increase in $[Ca^{2+}]_i$ has been reported[32] to lead to the opening of large-conductance Ca^{2+}-activated K^+ (BK) channels in astrocyte endfeet, releasing K^+ onto vessels. A caveat is that these experiments used a thromboxane analogue to maintain vessel tone in the brain slices studied, and this analogue has been shown to stimulate the trafficking of BK channels to the surface membrane of astrocytes[33]. It remains to be determined whether this K^+ release mechanism contributes to the regulation of blood flow *in vivo*.

There is strong evidence from brain slice and isolated retina preparations that astrocytes can control blood flow through the production and release of metabolites of arachidonic acid. When glutamate released from neurons activates astrocyte mGluRs (Fig. 2), the resultant rise in $[Ca^{2+}]_i$ activates phospholipase A_2, evoking the production of arachidonic acid from membrane phospholipids. The build-up of arachidonic acid leads, in turn, to the production of its metabolites, including prostaglandins and epoxyeicosatrienoic acids (EETs), which dilate nearby arterioles[15,34–37]. The prostaglandin involved is often suggested to be PGE_2, although many studies rely on inhibiting only the first enzyme (cyclooxygenase) in the prostaglandin-synthesis pathway, and other prostaglandins may contribute to the dilation. (The main pathways by which arachidonic acid derivatives are produced are shown in Fig. 3.) PGE_2 can relax vascular smooth muscle by binding to EP_4 prostaglandin receptors[38], which increase the activation of protein kinase A by cyclic AMP and thus decrease the

phosphorylation of the myosin light chain[39]. The dilation produced by PGE_2 and other arachidonic acid metabolites also partly reflects their activation of K^+ channels in vascular smooth muscle cells[40,41], making the membrane potential more negative and thus decreasing the entry of Ca^{2+} through voltage-gated channels. EETs may also elicit dilation by inhibiting receptors for thromboxane[42], a vasoconstricting derivative of arachidonic acid. *In vivo*, a rise in $[Ca^{2+}]_i$ in cortical astrocytes results in a vasodilation that is partly (~70%) mediated by prostaglandins[43]. The rest of the dilation could be mediated by EETs[36], probably derived from astrocytes[37], although this remains to be shown directly.

In brain slices and in the isolated retina rises in $[Ca^{2+}]_i$ in astrocytes can also constrict vessels[34,44]. This is mediated[34] by the conversion of arachidonic acid to 20-hydroxy-eicosatetraenoic acid (20-HETE), probably by the enzyme CYP4A in vascular smooth muscle cells (Fig. 2). Whether a rise in astrocyte $[Ca^{2+}]_i$ causes dilation or constriction may in part be determined by the pre-existing tone of the vessel[45], but the O_2 concentration also has a key role in determining this (see 'O$_2$ modulates neurovascular signalling'). *In vivo*, a rise in astrocyte $[Ca^{2+}]_i$ produces vasoconstriction[46], mediated by a phospholipase A_2 derivative (presumably 20-HETE), during cortical spreading depression — a pathological wave of cell depolarization and refractoriness lasting minutes, after a brief period of increased excitability, which spreads across the cortex.

Modulation of these arachidonic acid metabolic pathways for therapeutic effect will depend on understanding the isoforms of the enzymes that synthesize the vasoactive messengers. For example, prostaglandins that mediate physiological vasodilation in response to a rise in astrocyte $[Ca^{2+}]_i$ are mainly produced by cyclooxygenase 1 (COX1)[43], which is

expressed in astrocyte endfeet[35], with a possible contribution by COX3 (ref. 47) or COX2 (refs 35, 48). However, in pathological conditions, expression of COX2 is upregulated in astrocytes[48] and this might also contribute to prostaglandin synthesis. Neurons also express COX2 and PGE synthase, and this may account for a COX2 component of functional hyperaemia seen in the cortex[49].

Another, conceptually distinct, pathway by which astrocytes may control cerebral blood flow involves neurotransmitter uptake. Although about half of the increase in blood flow that is induced by neural activity in the olfactory bulb is mediated by glutamate activating astrocyte mGluRs and releasing prostaglandins, a further one-third of the increase is mediated by the activation of glutamate transporters on astrocytes[50]. The glutamate transporter component did not involve a $[Ca^{2+}]_i$ rise or prostaglandin release, and still occurred when glutamate receptors were blocked. In a related study in the visual cortex, blocking glutamate uptake prevented astrocyte $[Ca^{2+}]_i$ increases and light-reflectance changes that are attributed to increased blood flow[51]. However, interpretation of these data is complicated because, first, blood flow was not measured directly, second, it was unclear whether the astrocyte $[Ca^{2+}]_i$ rises were generated by mGluRs or by glutamate uptake, and third, blocking uptake will raise the extracellular glutamate concentration, which could desensitize glutamate receptors and have indirect effects. Similarly, in the olfactory bulb, co-transport of Na^+ with GABA taken up into astrocytes raises $[Na^+]_i$. This in turn inhibits Na^+/Ca^{2+} exchange, raising $[Ca^{2+}]_i$ and constricting arterioles[52], presumably by releasing arachidonic acid to generate 20-HETE (Fig. 2), although this was not tested.

The importance of astrocytes versus neurons

The relative importance of the neuronal and astrocytic vasodilating pathways remains a matter of debate, because synaptic glutamate release raises $[Ca^{2+}]_i$ in both cell types[53], with some astrocytes showing a fast rise in $[Ca^{2+}]_i$, similar to that in neurons[54]. Blood flow responses are strongly correlated with local field potentials that reflect synaptic and action potentials in neurons[55,56], and ionotropic glutamate receptor antagonists significantly decrease both field potentials and blood flow responses to stimulation in the cerebellum, cortex and olfactory bulb[16,17,55]. This might indicate that a rise of $[Ca^{2+}]_i$ in neuronal dendrites initiates much of the increase in blood flow[53]. However, the glutamate release that generates field potentials also activates astrocyte mGluRs, so field-potential amplitudes will also be correlated with astrocyte receptor activation.

A correlation of blood flow changes with field potentials is therefore not sufficient to prove that postsynaptic neuronal signalling dominates the control of blood flow. Furthermore, experiments blocking ionotropic glutamate receptors to prevent the release of neuronal messengers will decrease neuronal firing, inhibiting glutamate release onto astrocytes and thus preventing the release of astrocyte messengers. The potential for astrocytes to control blood flow is demonstrated by the fact that neural activity raises $[Ca^{2+}]$ in astrocytes[54,57], and that uncaging of Ca^{2+} in astrocytes dilates or constricts arterioles[34,35,43,44].

If the simplified scheme shown in Fig. 2 were correct, then the relative contributions of neuronal and glial signalling could be dissected by blocking either the neuronal NO pathway or the astrocyte arachidonic acid pathways. Knocking out or blocking nNOS (the only form of NOS contributing to functional hyperaemia[22]) does not affect whisker-stimulation-induced neural field potentials, but reduces the resulting blood flow increases by 37–60% in the somatosensory cortex[22,58]. In the cerebellum, knockout or inhibition of nNOS decreases activity-induced blood flow by 50–90%[13,24,25,59]. The remaining blood flow increase may be mediated by adenosine[14] or K^+ (ref. 60), but the possible contribution of prostaglandins and EETs has not been investigated. Blocking prostaglandin production with cyclooxygenase inhibitors reduces functional hyperaemia in the cortex[61,62] by 48–60% without affecting neuronal field potentials, whereas blocking EET production inhibits the response by 35–70%[36]. Thus, NO, cyclooxygenase and epoxygenase products are all important for generating functional hyperaemia in the cortex, whereas NO is more dominant in the cerebellum[59].

However, we will see below (see 'NO modulates astrocyte signalling') that the different pathways interact at the level of the enzymes producing arachidonic acid derivatives. In addition, a saturating interaction at the level of the arteriole smooth muscle membrane potential is expected if EETs and prostaglandins both dilate arterioles by opening K^+ channels, because once one pathway has hyperpolarized the cell out of the activation range of voltage-gated Ca^{2+} channels, then further opening of K^+ channels will have no effect. Caution is therefore needed in interpreting the results of experiments in which single enzymes are blocked.

Pathway–specific signalling differences

Even if functional hyperaemia is largely driven by synaptic glutamate release, it does not follow that all excitatory synapses are equally influential in controlling blood flow, even if they impinge on cells in the same
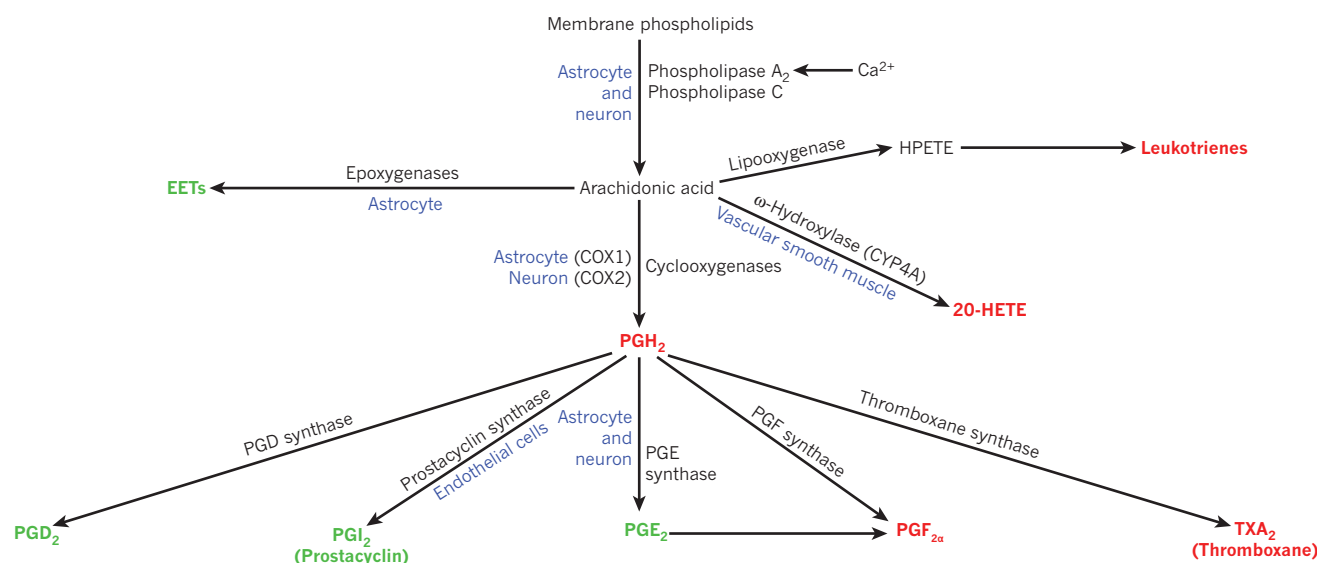
Figure 3 | Arachidonic acid metabolites that may contribute to control of cerebral blood flow. Arachidonic acid is formed from membrane phospholipids by Ca^{2+}-dependent and Ca^{2+}-independent lipases. Metabolites shown in green are vasodilators, red metabolites are vasoconstrictors, and blue denotes the location of some of the relevant enzymes. COX, cyclooxygenase; CYP, cytochrome P450 superfamily of enzymes; EET, epoxyeicosatrienoic acid; 20-HETE, 20-hydroxyeicosatetraenoic acid; HPETE, hydroperoxy-eicosatetraenoic acid.
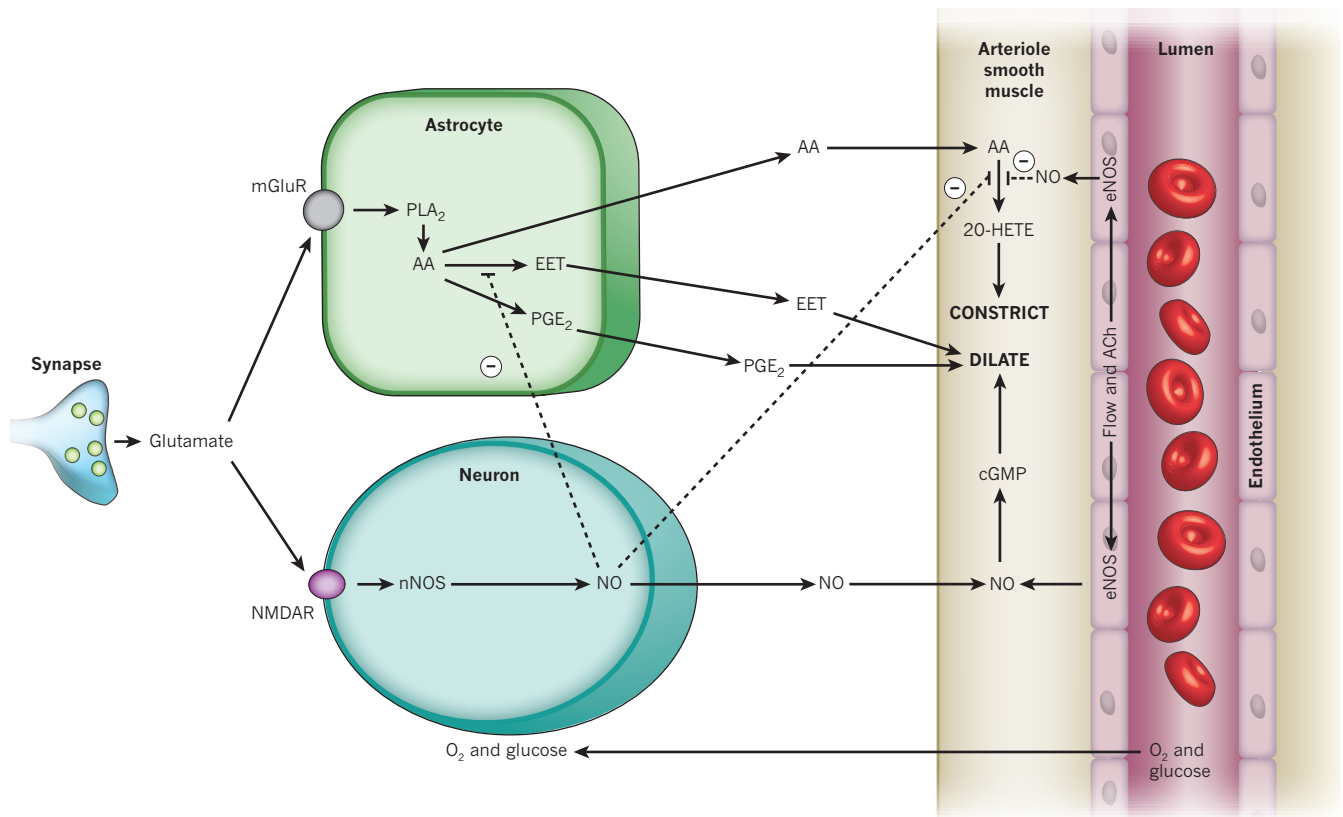
Figure 4 | Nitric oxide inhibits the production of key arachidonic acid-derived messengers. NO inhibits (dashed lines) the production of both the vasoconstricting 20-HETE and the vasodilating EETs[36,69]. NO also weakly stimulates COX1 and inhibits COX2 (not shown)[65]. Endothelial nitric oxide synthase (eNOS) can be activated by flow-induced shear stress or by acetylcholine (ACh). Other abbreviations as in earlier figures.

area. Indeed, activation of thalamocortical or transcallosal corticocortical inputs to the somatosensory cortex induces cerebral blood flow increases with a different dependence on the level of input activity[56], monotonically increasing with activity for the transcallosal pathway and showing a maximum at an intermediate level of input for the thalamocortical input. These differences may be partly due to differential activation of interneurons that release vasoactive peptides[56]. However, they may also reflect differences in the functional anatomy of neuro-vascular coupling for the two input pathways. For example, some of the 8,000 synapses innervating a typical rodent neocortical pyramidal cell[1] will release NO closer to arterioles, or induce a $[Ca^{2+}]_i$ rise in astrocytes that is more effective at releasing vasodilatory messengers onto arterioles, than will other synapses. Similarly, in the cerebellar cortex, although cerebral blood flow increases evoked by activation of the parallel and climbing fibre inputs to Purkinje cells depend partly on NO, the blood flow response to parallel fibre stimulation also involves a rise in $[K^+]_o$, whereas the response to climbing fibre activation partly involves adenosine release[60,63].

Thus, the relative importance of the neuronal and astrocyte pathways is likely to differ across brain areas, and even between different neural pathways in the same area. Consequently, the dependence of blood flow increases on the underlying neural activity will differ between different pathways, implying that functional imaging signals arising from these pathways will also reflect different aspects of neuronal function (see 'What does functional imaging measure?'). Similarly, therapeutic modulation of the signalling pathways controlling cerebral blood flow may have different effects in different brain areas, and on different neuronal pathways in the same area.

NO modulates astrocyte signalling

Assessing the relative significance of, or interfering therapeutically with, the different signalling pathways producing functional hyperaemia is complicated by interactions between them. The effects of the NO, cyclooxygenase and epoxygenase pathways can occlude each other. In the cortex, for example, functional hyperaemia is inhibited ~50% by blocking NOS and ~50% by blocking cyclooxygenase, but blocking both produces only a 70% inhibition[62]. Similarly, blocking NOS, blocking EET production, or blocking both pathways together all produce a ~60% inhibition of functional hyperaemia[37]. These results may be explained by interactions between the NO and arachidonic acid pathways, which are shown in Fig. 4.

Nitric oxide inhibits the enzymes that synthesize 20-HETE and EETs[64]. It also, much more weakly, inhibits prostaglandin production by COX2, but stimulates COX1 (ref. 65). The inhibition of the synthesis of the vasoconstrictor 20-HETE by NO is of particular interest for two reasons. First, this effect may underlie a significant fraction of the dilating effect of NO[66] (20–60% in different vessels[64]), which is independent of the NO vasodilation mediated by cyclic GMP (Fig. 4). Second, it provides a probable explanation for why, in the neocortex, NO is required for functional hyperaemia to occur, but is not the primary mediator of the increase in blood flow[23]. If the main cause of the blood flow increase is the generation of arachidonic acid by astrocytes, then having NO present to inhibit 20-HETE formation will ensure that only the vasodilatory prostaglandin and EET derivatives of the arachidonic acid will affect arteriole diameter. In the retina, however, where arteriole vasodilations are produced by EETs[34], the presence of NO promotes light-induced constrictions and inhibits dilations, suggesting that the production of EETs is more NO sensitive than is the production of 20-HETE[34].

O₂ modulates neurovascular signalling

Variations in O_2 concentration in brain tissue alter neurovascular coupling in two ways. The O_2 level affects the synthesis of the glial and neuronal messengers involved, and also alters the levels of lactate and adenosine that modulate the pathways by which these messengers
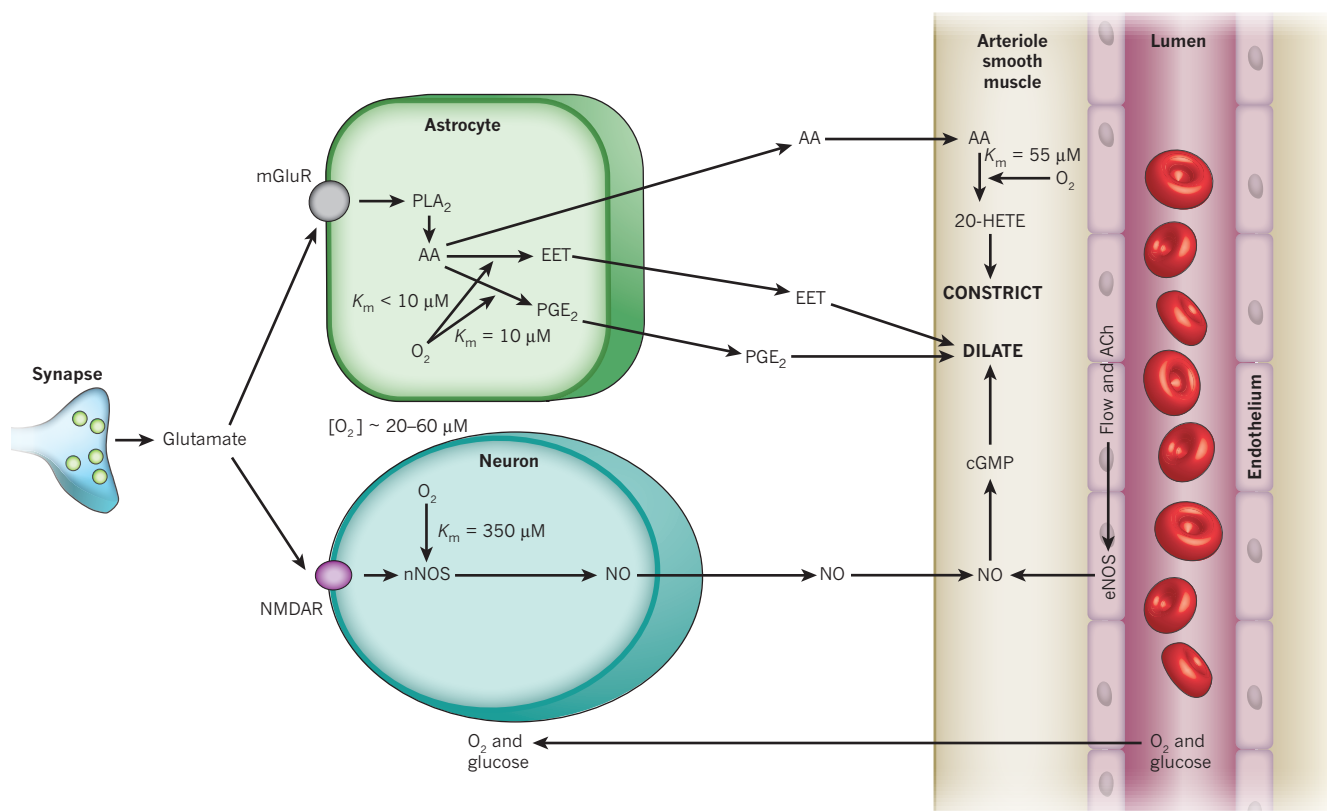
Figure 5 | Oxygen differentially affects the synthesis of neurovascular messengers. The concentration of O_2 in the extracellular space is 20–60 μM. This is significantly higher than the effective K_m for O_2 activating the enzymes synthesizing EETs and prostaglandins, but is in a range in which changes in O_2 concentration will modulate the production of NO and 20-HETE. Abbreviations as in earlier figures.

regulate vascular tone.

O_2 is needed for the synthesis of nitric oxide and the vasoactive messengers derived from arachidonic acid. Comparing *in situ* O_2 levels with the O_2 affinities of the O_2-sensitive reactions (Fig. 5) suggests how changes in O_2 concentration will affect neurovascular coupling. As O_2 concentrations are decreased, the synthesis of NO by neurons is expected to be inhibited first (Michaelis constant (K_m) ~ 350 μM at 25 °C[67]), followed by 20-HETE synthesis (K_m ~ 55 μM at 37 °C[68]), whereas EET synthesis (K_m < 10 μM at 37 °C[68]) and prostaglandin synthesis (K_m ~ 10 μM at 24 °C[69]) should be maintained at much lower O_2 concentrations. How do these values compare with the O_2 levels found in brain tissue? Neurophysiologists often use 95–100% O_2 (~1 mM, or 760 mm Hg) in solutions superperfusing isolated tissue, which lead to O_2 concentrations of ~100–150 μM in brain slices. This is much higher than the *in vivo* values[18,19] of 20–60 μM, which can even decrease by a further 13 μM during intense synaptic activity[18,19]. These *in vivo* values, which can be mimicked by superfusing brain slices with solution equilibrated with 20% O_2 (ref. 70), are comparable to or lower than the K_m values for O_2 in the synthesis of 20-HETE and NO. Thus, at *in vivo* levels of O_2, we expect NO and 20-HETE synthesis to be significantly limited by the amount of O_2 available.

Consistent with this view, lowering the O_2 concentration in solution superfusing brain slices from 95% to 20% (to lower the tissue [O_2] from about 125 to 40 μM) has a dramatic effect on the vascular responses produced by uncaging Ca^{2+} in astrocytes[35]. With a supraphysiological O_2 concentration in the tissue, raising astrocyte [Ca^{2+}]$_i$ led to arteriolar constriction, whereas with a physiological [O_2], vasodilation occurred. These changes are, in part, predicted from the modulatory effects of O_2 and NO on the signalling pathways outlined above (in this and the preceding section). The suppression of 20-HETE formation by the lower O_2 concentration is expected to reduce the vasoconstriction produced when arachidonic acid is generated in astrocytes (Fig. 5), and the lower

[O_2] will also result in less NO being present to inhibit the formation of vasodilatory EETs (Fig. 4) in tissues where these contribute to dilation.

However, two other factors were also shown to contribute significantly to the effects of lowering [O_2] (ref. 35; Fig. 6). As O_2 concentrations decrease, the lack of energy for ATP synthesis causes an increase in the level of extracellular adenosine (Fig. 6), which binds to adenosine A_{2A} receptors on vascular smooth muscle cells to depress vessel constriction. In addition, a decrease in the rate of oxidative phosphorylation relative to the rate of glycolysis results in lactate production (Fig. 6). Monocarboxylate transporters release the lactate into the extracellular space, where it reduces the clearance of extracellular PGE_2 by the prostaglandin transporter (Fig. 6). Thus, when PGE_2 is released from astrocytes, extracellular PGE_2 increases to a greater degree, resulting in larger arteriole dilations[35]. This effect of lactate may partly explain why, in humans and rats *in vivo*, cerebral blood flow is regulated by the lactate/pyruvate concentration ratio and thus by the NADH/NAD$^+$ ratio[12]. Interestingly, lactate is released into the extracellular space during synaptic activity[71], which should promote vasodilation.

Despite our understanding of how O_2 levels regulate astrocyte-mediated neurovascular coupling in brain slices[35], imposing artificially high O_2 concentrations *in vivo* does not lead to smaller vasodilations or the emergence of vasoconstrictions[7]. This could reflect a potentiation of NO-mediated signalling from neurons to arterioles (relative to that occurring in brain slices superfused with 95% O_2) by the high O_2 level produced by hyperbaric O_2 (ref. 7) (as NOS activity is potentiated by the higher [O_2]; Fig. 5), which outweighs the effect of O_2 on astrocyte-mediated signalling.

Control of blood flow at the capillary level

Until recently, it was assumed that neurovascular coupling is mediated solely by changes in the tone of the smooth muscle that forms a continuous layer around arterioles. This idea has been challenged by
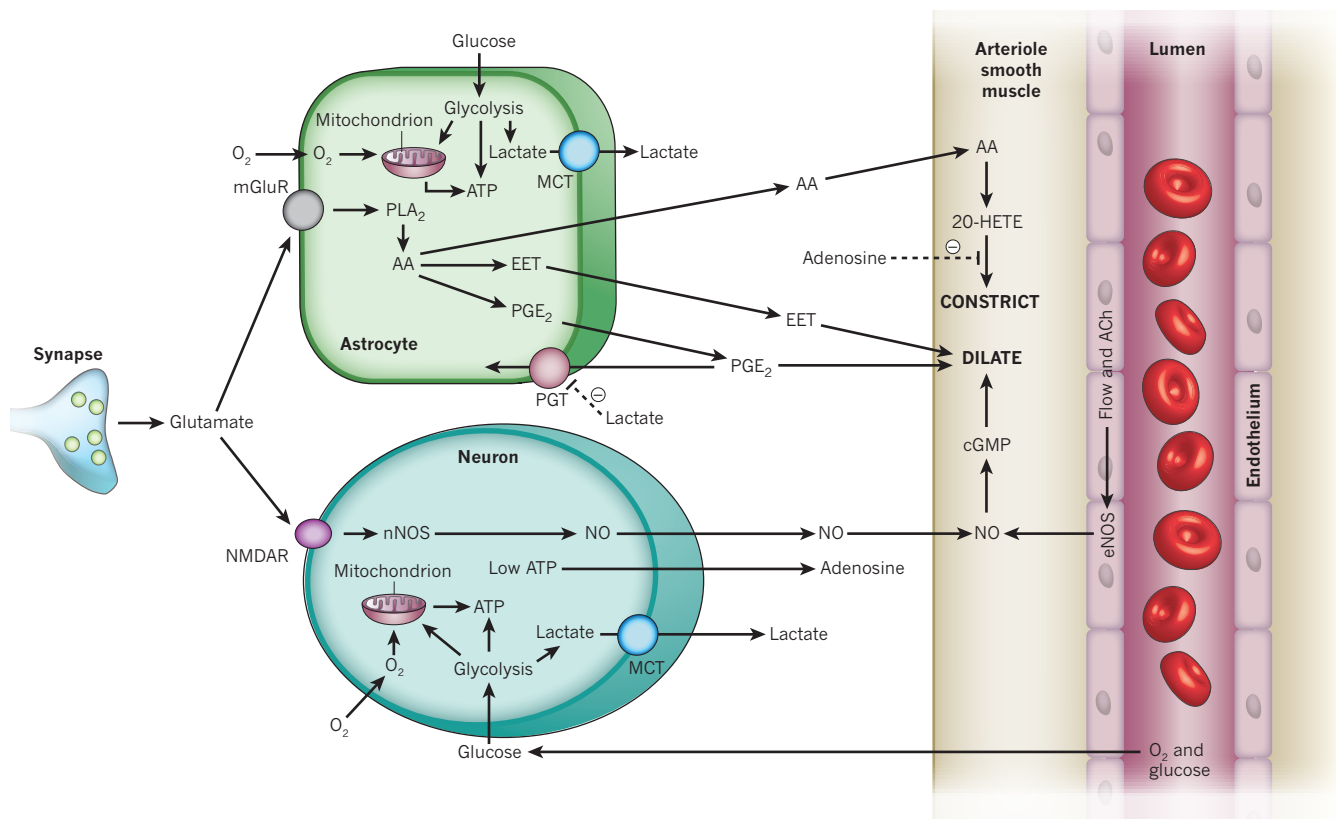
Figure 6 | Lactate and adenosine affect neurovascular signalling at low [O₂]. Low O_2 concentrations lead to mitochondrial oxidative phosphorylation failing to consume all the pyruvate produced by glycolysis, resulting in an export of lactate by monocarboxylate transporters (MCTs). Extracellular lactate inhibits the reuptake of PGE_2 by the prostaglandin transporter (PGT), promoting vasodilation[35]. Low energy levels also lead to the formation of adenosine, which inhibits 20-HETE-mediated arteriolar constriction by acting on adenosine A_{2A} receptors[35]. Other abbreviations as in earlier figures.

the discovery that pericytes — cells present at ~50-µm intervals along capillaries — can markedly alter capillary diameter, and thus potentially regulate cerebral blood flow at the capillary level.

Pericytes express contractile proteins, and their location on capillaries led to the suggestion, some 130 years ago, that they could constrict the microvasculature[72]. This idea was reinforced with observations that substances that alter arteriole diameter, including arachidonic acid derivatives and neurotransmitters, can contract and relax pericytes cultured on rubber membranes[73]. Importantly, in a series of papers, pericytes on isolated retinal capillaries were demonstrated to constrict or dilate in response to neurotransmitters, as a result of $[Ca^{2+}]_i$ alterations[74]. It is assumed that the signalling pathways controlling pericyte constriction and dilation will be similar to those for arterioles shown in Fig. 2, but this still needs to be tested. *In situ* in brain slices, pericytes constrict in response to noradrenaline and dilate in response to glutamate[75], and in the isolated retina, blocking ionotropic GABA receptors constricts capillaries, demonstrating that endogenous transmitter release can regulate capillary diameter[75]. Regulation of cerebral blood flow at the capillary level has yet to be demonstrated *in vivo* in physiological conditions[72], although it does occur after ischaemia[76].

Signals for contraction (presumed to be depolarization) and perhaps for dilation (hyperpolarization) can propagate from one pericyte to another[74,75]. This signal spread may occur through gap junctions between the interdigitating processes of the pericytes themselves, or through gap junctions with endothelial cells. Because active neurons are, on average, closer to pericytes than to arterioles[77] (8–23 µm away versus 70–160 µm), this signal spread raises the theoretical possibility that vascular responses to changes in neuronal activity may be initiated by pericytes and then propagated to upstream arterioles.

The importance of capillary diameter control by pericytes for regulating cerebral blood flow will depend on the fraction of the haemodynamic resistance of the vascular network that the capillaries contribute.

Models in the literature[78,79] suggest that the capillary resistance is between 16% and 70% of the total vascular resistance (arterioles plus capillaries plus venules) within the brain parenchyma. If blood flow in capillaries obeys Poiseuille's law of fluid dynamics (see below), with flow proportional to the fourth power of the vessel diameter, then these values lead to the conclusion that a 2.1-fold capillary dilation produced by superfused glutamate in the presence of noradrenaline[75] could, in principle, increase blood flow 1.18- to 2.98-fold[72], and thus contribute significantly to functional hyperaemia. However, this may overestimate the role of pericytes if endogenous noradrenaline release is less than that applied when probing the effect of glutamate[75], or if the effect of superfused glutamate in brain slices overestimates dilations produced in perfused vessels *in vivo* by neurons releasing glutamate. On the other hand, Poiseuille's law severely underestimates the power of pericytes to regulate blood flow, because red and white blood cells have to deform considerably to pass through capillaries that constrict below ~5 µm — or they may not be able to pass through at all, which becomes important when pericytes constrict capillaries after ischaemia[76]. This will result in a more non-linear dependence of flow on diameter, enhancing the influence of capillaries on blood flow. Experiments are needed to define the importance of pericytes for regulating blood flow *in vivo*.

Blood flow rises more than energy use

Traditional analyses of sensory systems assumed that external stimuli induce action potentials in neurons with receptive fields 'tuned' to particular features of the stimuli. On this basis, low levels of spontaneous neuronal activity merely provide a 'noise' background against which the stimulus must be detected, and sensory input should greatly increase neuronal activity and hence cortical energy use. Recently, however, it has been suggested that incoming sensory information produces only small changes to the ongoing activity level in cortical neurons[80]. Indeed, across a range of sensory systems, less than a 10% change of neuronal

spiking is produced by perceptual tasks used in functional imaging experiments[81], and as a result sensory input alters cortical energy use by a relatively small fraction[82].

The fractional increase in blood flow induced by sustained neuronal activity is at least 4-fold greater than the increase in ATP consumption by the neurons[82]. This is consistent with blood flow being mainly regulated by feedforward neurotransmitter-mediated mechanisms rather than by a negative-feedback loop driven by energy demand (Fig. 1b, c), because a negative-feedback system could not produce a sustained increase in energy supply that is larger than the increase in energy consumption. As explained below (see 'What does functional imaging measure?'), this disproportionate increase in blood flow is the basis of the blood-oxygen-level dependent (BOLD) functional imaging technique. It was initially thought that most of the energy used when neurons are activated came from glycolysis rather than oxidative phosphorylation, because neuronal activity increases glucose uptake much more than O_2 uptake[83]. However, later studies found less of a difference in the glucose and O_2 uptake rates[84], and far more ATP is produced by oxidative phosphorylation than by glycolysis per glucose molecule. As a result, even with a much smaller increase in O_2 uptake than in glucose uptake[83], at least 60% of the extra ATP production would be occurring by oxidative phosphorylation[85], and more recent work suggests that almost all of the ATP is generated by oxidative metabolism of glucose[82].

A possible explanation for the large increase in blood flow induced by activity, relative to the increase in O_2 uptake and ATP production, was based on modelling of O_2 uptake: a large increase in blood flow (driven by glutamate release, not by O_2 deficit) may be needed to generate even a small increase in O_2 uptake[86]. However, pharmacologically blocking most of the increase in blood flow produced by activity has little effect on the increase in O_2 consumption[87], arguing against this explanation. At present, therefore, it seems that a large increase in blood flow (and glucose uptake) should not be needed to maintain energy supply during normal brain activity. Perhaps this large increase occurs as a by-product of a system that attempts to maintain blood flow during conditions of greater energy need that can occur pathologically (see 'Spreading depression and neurovascular coupling').

What does functional imaging measure?

BOLD functional magnetic resonance imaging (fMRI) of brain activity depends crucially on functional hyperaemia[88]. This technique detects a magnetic resonance radio signal emitted from proton spins. The strength of the signal is decreased by deoxyhaemoglobin, which is paramagnetic and makes the magnetic field less uniform. Without functional hyperaemia, the O_2 consumption powering neuronal activity would increase the deoxyhaemoglobin level and thus decrease the MRI signal. In fact, BOLD images show an increase in intensity in active brain areas. This is because the increase in blood flow evoked by neuronal activity brings in excess oxygenated blood, so that the deoxyhaemoglobin level falls and the MRI signal is increased. Thus, simplistically, the size of the BOLD fMRI signal is determined by the difference between the amount of functional hyperaemia, which increases the signal, and the use of O_2 by neurons, which reduces the signal (there is also a dependence on blood volume, which increases with the pressure increase driving the increased flow, which is ignored for simplicity here).

Because the effect of functional hyperaemia dominates, it follows that, to a large extent, BOLD fMRI signals reflect the causes of functional hyperaemia. When it was thought that activity-evoked blood flow increases were driven by energy use, fMRI images were believed to reflect energy use by active neurons. The discovery that functional hyperaemia is driven largely by glutamate release indicates, instead, that fMRI mainly images the neurovascular signalling consequences of synaptic activity[89]. Consistent with this, cortical BOLD signals correlate slightly better with field potentials reflecting stimulus-evoked synaptic currents than with stimulus-evoked action potentials[90], but they still correlate well with action potential activity[90] because cortical synaptic potentials are themselves well correlated with pyramidal cell spiking.

It would be instructive to repeat this experiment in the cerebellum, where parallel fibre activity leads to a glutamate-mediated increase in blood flow, but to a decrease in principal (Purkinje) cell firing due to disynaptic inhibition[55].

Because BOLD signals mainly reflect the causes of functional hyperaemia, if we assume that this is largely produced by astrocyte signalling, BOLD fMRI images will essentially reflect the activation of astrocyte mGluRs (Fig. 2). Conversely, if the neuronal NO pathway dominates, then BOLD essentially images the activation of postsynaptic nNOS. But BOLD may still give information on neuronal spiking activity, for two reasons. First, glial $[Ca^{2+}]_i$ increases (and downstream generation of arachidonic acid derivatives) can be closely correlated with the activity of adjacent neurons[51], presumably because the glia respond to glutamate released at synapses onto the neurons. Second, typical voxel sizes used in fMRI (~1.5–3 mm) include many output synapses from the local axon collaterals of pyramidal cells in the cortex[91], and so the BOLD signal obtained, although originating from functional hyperaemia evoked by glutamate release, could nevertheless reflect the local pyramidal cell spiking activity that drives glutamate release at those synapses. For this to be the case, the effect on the vasculature of the axon collateral synapses would have to outweigh the effect of synapses carrying input to that part of the cortex from other brain areas.

The spatial resolution of BOLD fMRI has increased with the field strength used for imaging, and now approaches 200 μm in animals[85]. Even higher resolution can be obtained by optical imaging of changes in the oxy- and deoxyhaemoglobin concentrations within vessels[92]. The different spatial ranges of the vasodilatory mechanisms outlined in Fig. 2, that is, the distance that NO[93] or arachidonic acid derivatives can diffuse and the distance that $[Ca^{2+}]_i$ increases propagate through astrocytes, will limit the spatial resolution that functional imaging can potentially attain.

Spreading depression and neurovascular coupling

Functional hyperaemia could be particularly important for preventing neuronal death in pathological conditions in which energy use is raised, but it is in precisely these conditions that neurovascular coupling sometimes fails. One example is in cortical spreading depression, which occurs in migraine[94], when it is relatively benign, and repeatedly after stroke, subarachnoid haemorrhage or brain trauma, when it is associated with delayed neurological deficits[95–97]. Spreading depression is associated with a notable failure of brain ion homeostasis[98]. The extracellular potassium concentration rises transiently from ~3 to ~50 mM, depolarizing neurons and astrocytes, and the extracellular glutamate concentration rises[99] as a result of release from neurons and impairment of uptake[100]. Brain energy use increases[101], presumably because the increase in glutamate concentration activates an influx of Na^+ and Ca^{2+} ions that need to be pumped out, and the sodium pump is further activated by the rise in $[K^+]_o$. As a result, the cortex consumes more O_2 and glucose than is provided by the blood, resulting in a decrease in their extracellular concentrations[101–103]. This increase in energy use is pronounced during the first few minutes of spreading depression, but persists at a lower level for hours[101].

A failure of neurovascular coupling to provide a sufficient increase in blood flow during the increased energy consumption generated by spreading depression may underlie the neurological deficits associated with spreading depression after subarachnoid haemorrhage[97]. During spreading depression, cerebral blood flow is initially transiently increased[101,102] (less so in migraine than in stroke or brain trauma[94,104]), sometimes with a preceding decrease in blood flow[46,105]. After a few minutes, however, blood flow is reduced[94,101,102,105] by 20–30%. This decrease lasts for hours despite the increase in energy consumption during this period[101].

The rise in $[K^+]_o$ and associated cell depolarization during spreading depression raises $[Ca^{2+}]_i$ in neurons and astrocytes[46], leading to the release of messengers that evoke the observed blood flow changes. A release of calcitonin gene related peptide (CGRP) from trigeminal neurons innervating cerebral arteries, probably as a result of synaptic terminals depolarizing, contributes to the transient increase in blood

flow induced by spreading depression[104,106]. CGRP may act on receptors on smooth muscle cells or on astrocytes to produce this effect, although blocking the production of prostaglandins or EETs does not reduce the blood flow increase[104], indicating that a direct action on smooth muscle is more likely. Release of NO has also been suggested to contribute to the vasodilation[105,106], although this is controversial[104] and the effects of NOS blockers may reflect a need for NO to be present to suppress vasoconstriction mediated by 20-HETE (Fig. 4).

The sustained decrease in blood flow caused by spreading depression may be partly mediated by 20-HETE generated from astrocyte arachidonic acid. Reducing the increase in astrocyte $[Ca^{2+}]_i$ that accompanies spreading depression (using thapsigargin to prevent Ca^{2+} accumulation in internal stores), or blocking phospholipase A_2 to prevent arachidonic acid formation, both prevented the vasoconstriction[46]. This study, however, imaged only the early vasoconstriction occurring before the transient increase of blood flow, and not the long-lasting constriction that reduces blood flow for hours. During the prolonged constriction, normal functional hyperaemia is disrupted: vessels no longer dilate to physiological activation[107]. This is partly due to an impairment of the NO system[105,108], which normally dilates vessels directly or suppresses 20-HETE-mediated constrictions (Fig. 4), and may reflect the fall in O_2 concentration from ~24 to ~12 µM during this period[101], which will approximately halve NO production (Fig. 5). Indeed, more generally, the fact that O_2 supply by the blood depends partly on the release of the vasodilator NO, which requires O_2 for its formation, suggests a potentially damaging feedback loop. An increase in energy consumption lowers the O_2 concentration[101], which decreases NO formation, lowering blood flow and reducing the O_2 concentration further.

Neurovascular coupling after ischaemia

Following brain ischaemia, when an occluding intraluminal thrombus is cleared from a blood vessel either spontaneously or therapeutically using exogenous tissue plasminogen activator (tPA), after a brief period of hyperaemia there is a decrease in blood flow lasting several hours[2,109–111]. This inadequate matching of blood flow to neural activity may produce damage to neurons or glial cells beyond that caused by the initial ischaemia. The decreased flow, or 'no-reflow phenomenon', was attributed to a reduction in capillary diameter as a result of astrocyte endfoot swelling, causing capillary blockage by blood cells and fibrin[109], but later work suggested that this effect had been overestimated[112]. A failure of arteriole vasodilation mechanisms (tested with CO_2 and acetylcholine, which both release NO[113] and arachidonic acid derivatives[114]) may underlie the long-lasting decrease in blood flow[2,110,115]. However, in studies measuring blood flow and not vessel diameter, defects in arteriole dilation would be hard to distinguish from a situation in which arteriole dilation occurs but fails to increase blood flow because some capillaries are blocked.

Recent evidence indicates an important role for capillary pericytes in the long-lasting blood flow decrease following ischaemia. Some pericytes constrict at the start of ischaemia[75]. This may be because their $[Ca^{2+}]_i$ rises in the absence of ATP to pump Ca^{2+} out of the cell, although altered release of vasoactive messengers may also contribute (see below). Indeed, a decreased capillary diameter has been noted after ischaemia[111]. Interestingly, pericytes are also highly susceptible to damage in ischaemia[76], when ATP levels are expected to be low. This raises the possibility that pericytes constrict capillaries at the start of a stroke, and then stay in rigor (because no ATP is available to relax their contractile filaments), causing the capillaries to remain too small for the passage of red blood cells. In agreement with this, pericytes remain constricted and prevent red blood cell passage even two hours after re-opening an occluded parent artery[76]. Suppression of oxidative and nitrosative stress prevents this pericyte constriction, restores the patency of capillaries, and improves tissue recovery[76]. The involvement of oxidative or nitrosative stress suggests another mechanism by which pericytes might constrict. Peroxynitrite inhibits the formation of vasodilating prostacyclin by endothelial cells[116], resulting in the accumulation of vasoconstricting PGH_2 (Fig. 3),

and may also inhibit the formation of vasodilatory EETs[117] and NO[118], any of which could lead to pericyte constriction.

The therapeutic use of tPA may itself alter neurovascular coupling, in part because it may extravasate and gain access to the brain parenchyma when the blood–brain barrier becomes leaky in ischaemia (a process that may be enhanced by tPA upregulating[119] matrix metalloprotease 9). There is debate about whether tPA is neurotoxic[119], but endogenous tPA, which is released from neurons, cleaves the NR1 subunit of NMDA receptors and thus enhances NMDA receptor signalling[120]. Independent of this effect, endogenous tPA is essential for neurovascular coupling mediated by NMDA receptors and NO, because it regulates the phosphorylation of nNOS and thereby promotes NO release[121]. Accordingly, exogenous tPA promotes vessel dilation in an NO-dependent manner[122], but it also enhances the failure of vasodilation that occurs after stroke[123]. To complicate the interpretation of these data, when used clinically tPA is co-packaged with a vehicle, a high concentration of L-arginine, which will promote NO formation by NOS. This will contribute to vasodilation and could possibly damage neurons.

Neurovascular coupling in Alzheimer's disease

There is increasing evidence for vascular factors having a causal role in the development of Alzheimer's disease[3]. Many patients with Alzheimer's disease have regional cerebral hypoperfusion[124], which correlates with cognitive decline[125]. Amyloid-β peptide decreases functional hyperaemia by promoting oxidative stress[126], which inhibits the production of astrocytic and neuronal vasodilating messengers[116–118]. A raised basal $[Ca^{2+}]_i$ and enhanced occurrence of spontaneous Ca^{2+} waves in the astrocytes of mice with amyloid-β plaques[127], together with increased contractility of vascular smooth muscle cells[128], will also disrupt the normal regulation of cerebral blood vessels by astrocytes and neurons. Such changes in neurovascular coupling may make it hard to interpret BOLD fMRI signals from older subjects[129], just as inhibiting neurovascular coupling pharmacologically alters the relationship between neuronal activity and BOLD signals[20]. Similarly, the formation of an astrocytic scar after traumatic brain injury is likely to profoundly alter neurovascular coupling.

Prospects for new therapies

There have been four major conceptual shifts in our understanding of how cerebral blood flow is regulated. It is now thought that neurotransmitters, particularly glutamate, rather than energy use are the principal agents generating activity-induced blood flow; that astrocytes mediate a large part of this blood flow control; that O_2 concentration regulates the relative importance of the signalling pathways involved; and that control of blood flow occurs at the capillary level as well as the arteriole level. These developments not only provide an increased understanding of how neural computation is powered, they also offer opportunities for developing therapies for treating disorders of cerebral blood flow.

In general, identifying the signalling pathways that regulate cerebral blood flow provides opportunities for manipulating those pathways therapeutically. This is conceptually straightforward for neurotransmitter-mediated signalling pathways that depend on well understood receptors and enzymes that can, in principle, be blocked. This could be done, for example, at the level of astrocyte mGluRs or the enzymes that generate vasoactive messengers such as cyclooxygenase, epoxygenase, CYP4A or NOS. The existence of many pathways regulating blood flow (Fig. 2) may allow subtle therapeutic modification of blood flow in different conditions or brain areas, so long as interactions between the different pathways (Fig. 4) are taken into account. The advent of new imaging techniques for rapid assessment of patients' cerebral blood flow could also enable therapies to be given earlier, before neuronal damage has been initiated. In particular, the signalling changes that occur in spreading depression and stroke suggest new therapeutic approaches, as follows.

The decrease in blood flow that follows spreading depression[101,102,105], which may contribute to cognitive decline after brain trauma, stroke

or vasospasm after subarachnoid haemorrhage[97], may be produced by excessive generation of vasoconstricting 20-HETE formed from astrocyte arachidonic acid. This could be tested using inhibitors of the CYP4A enzyme that generates 20-HETE. A role for pericyte constriction in this phenomenon should also be tested.

The long-lasting decrease in blood flow observed after experimental ischaemia[2,109–111], produced in part by pericyte constriction[75,76], indicates that even rapid dissolution of a clot using tPA may be limited in its effects on restoring perfusion. This suggests that, in combination with tPA application, it would be worth trying to preserve normal pericyte function. Pericyte constriction and death after ischaemia are mediated by oxidative or nitrosative stress[76], so infusing antioxidants at the same time as tPA may be a useful adjuvant approach. Preclinical studies assessing how pericyte properties vary with age and gender, and whether pericyte contraction contributes to pathology in hypertension (which is associated with neurovascular dysfunction[3]) or hyperglycaemia[72,74,130], will also be valuable. ∎

1. Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21,** 1133–1145 (2001).
2. Leffler, C. W., Busija, D. W., Mirro, R., Armstead, W. M. & Beasley, D. G. Effects of ischemia on brain blood flow and oxygen consumption of newborn pigs. *Am. J. Physiol.* **257,** H1917–H1926 (1989).
3. Girouard, H. & Iadecola, C. Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease. *J. Appl. Physiol.* **100,** 328–335 (2006).
4. Baptiste, D. C. & Fehlings, M. Pharmacological approaches to repair the injured spinal cord. *J. Neurotrauma* **23,** 318–334 (2006).
5. Tian, R. *et al.* Role of extracellular and intracellular acidosis for hypercapnia-induced inhibition of tension of isolated rat cerebral arteries. *Circ. Res.* **76,** 269–275 (1995).
6. Mintun, M. A. *et al.* Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. *Proc. Natl Acad. Sci USA* **98,** 6859–6864 (2001).
7. Lindauer, U. *et al.* Neurovascular coupling in rat brain operates independent of hemoglobin deoxygenation. *J. Cereb. Blood Flow Metab.* **30,** 757–768 (2010). **Challenges brain slice data showing that high [O₂] converts dilations seen at physiological [O₂] into constrictions.**
8. Powers, W. J., Hirsch, I. B. & Cryer, P. E. Effect of stepped hypoglycemia on regional cerebral blood flow response to physiological brain activation. *Am. J. Physiol.* **270,** H554–H559 (1996).
9. Astrup, J. *et al.* Evidence against H⁺ and K⁺ as main factors for the control of cerebral blood flow: a microelectrode study. *Ciba Found. Symp.* **56,** 313–337 (1978).
10. Makani, S. & Chesler, M. Rapid rise of extracellular pH evoked by neural activity is generated by the plasma membrane calcium ATPase. *J. Neurophysiol.* **103,** 667–676 (2010).
11. Ko, K. R., Ngai, A. C. & Winn, H. R. Role of adenosine in regulation of regional cerebral blood flow in sensory cortex. *Am. J. Physiol. Heart Circ. Physiol.* **259,** H1703–H1708 (1990).
12. Ido, Y., Chang, K., Woolsey, T. A. & Williamson, J. R. NADH: sensor of blood flow need in brain, muscle and other tissues. *FASEB J.* **15,** 1419–1421 (2001).
13. Akgören, N., Fabricius, M. & Lauritzen, M. Importance of nitric oxide for local increases of blood flow in rat cerebellar cortex during electrical stimulation. *Proc. Natl Acad. Sci. USA* **91,** 5903–5907 (1994).
14. Li, J. & Iadecola, C. Nitric oxide and adenosine mediate vasodilation during functional activation in cerebellar cortex. *Neuropharmacology* **33,** 1453–1461 (1994).
15. Zonta, M. *et al.* Neuron-to-astrocyte signaling is central to the dynamic control of brain microcirculation. *Nature Neurosci.* **6,** 43–50 (2003).
16. Nielsen, A. N. & Lauritzen, M. Coupling and uncoupling of activity-dependent increases of neuronal activity and blood flow in rat somatosensory cortex. *J. Physiol.* **533,** 773–785 (2001).
17. Chaigneau, E. *et al.* The relationship between blood flow and neuronal activity in the rodent olfactory bulb. *J. Neurosci.* **27,** 6452–6460 (2007).
18. Offenhauser, N., Thomsen, K., Caesar, K. & Lauritzen, M. Activity induced tissue oxygenation changes in rat cerebellar cortex: interplay of postsynaptic activation and blood flow. *J. Physiol.* **565,** 279–294 (2005).
19. Lecoq, J. *et al.* Odor-evoked oxygen consumption by action potential and synaptic transmission in the olfactory bulb. *J. Neurosci.* **29,** 1424–1433 (2009).
20. St Lawrence, K. S., Ye, F. Q., Lewis, B. K., Frank, J. A. & McLaughlin, A. C. Measuring the effects of indomethacin on changes in cerebral oxidative metabolism and cerebral blood flow during sensorimotor activation. *Magn. Reson. Med.* **50,** 99–106 (2003).
21. Busija, D. W., Bari, F., Domoki, F. & Louis, T. Mechanisms involved in the cerebrovascular dilator effects of *N*-methyl-D-aspartate in cerebral cortex. *Brain Res. Rev.* **56,** 89–100 (2007).
22. Ma, J., Ayata, C., Huang, P. L., Fishman, M. C. & Moskowitz, M. A. Regional cerebral blood flow response to vibrissal stimulation in mice lacking type I NOS gene expression. *Am. J. Physiol.* **270,** H1085–H1090 (1996).
23. Lindauer, U., Megow, D., Matsuda, H. & Dirnagl, U. Nitric oxide: a modulator, but not a mediator, of neurovascular coupling in rat somatosensory cortex. *Am. J. Physiol. Heart Circ. Physiol.* **277,** H799–H811 (1999).
24. Akgören, N., Dalgaard, P. & Lauritzen, M. Cerebral blood flow increases evoked by electrical stimulation of rat cerebellar cortex: relation to excitatory synaptic activity and nitric oxide synthesis. *Brain Res.* **710,** 204–214 (1996).
25. Yang, G., Zhang, Y., Ross, M. E. & Iadecola, C. Attenuation of activity-induced increases in cerebellar blood flow in mice lacking neuronal nitric oxide synthase. *Am. J. Physiol. Heart Circ. Physiol.* **285,** H298–H304 (2003).
26. Cauli, B. *et al.* Cortical GABA interneurons in neurovascular coupling: relays for subcortical vasoactive pathways. *J. Neurosci.* **24,** 8940–8949 (2004).
27. Kocharyan, A., Fernandes, P., Tong, X. K., Vaucher, E. & Hamel, E. Specific subtypes of cortical GABA interneurons contribute to the neurovascular coupling response to basal forebrain stimulation. *J. Cereb. Blood Flow Metab.* **28,** 221–231 (2008).
28. Knot, H. J., Zimmermann, P. A. & Nelson, M. T. Extracellular K⁺-induced hyperpolarizations and dilations of rat coronary and cerebral arteries involve inward rectifier K⁺ channels. *J. Physiol.* **492,** 419–430 (1996).
29. Paulson, O. B. & Newman, E. A. Does the release of potassium from astrocyte endfeet regulate cerebral blood flow? *Science* **237,** 896–898 (1987).
30. Metea, M. R., Kofuji, P. & Newman, E. A. Neurovascular coupling is not mediated by potassium siphoning from glial cells. *J. Neurosci.* **27,** 2468–2471 (2007).
31. Porter, J. T. & McCarthy, K. D. Hippocampal astrocytes *in situ* respond to glutamate released from synaptic terminals. *J. Neurosci.* **16,** 5073–5081 (1996).
32. Filosa, J. A. *et al.* Local potassium signaling couples neuronal activity to vasodilation in the brain. *Nature Neurosci.* **9,** 1397–1403 (2006).
33. Ou, J. W. *et al.* Ca²⁺- and thromboxane-dependent distribution of MaxiK channels in cultured astrocytes: from microtubules to the plasma membrane. *Glia* **57,** 1280–1295 (2009).
34. Metea, M. R. & Newman, E. A. Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. *J. Neurosci.* **26,** 2862–2870 (2006).
35. Gordon, G. R. J. *et al.* Brain metabolism dictates the polarity of astrocyte control over arterioles. *Nature* **456,** 745–749 (2008). **Shows that O₂ level profoundly affects vascular response to neuronal activity.**
36. Peng, X. *et al.* Suppression of functional hyperemia to vibrissal stimulation in the rat by epoxygenase inhibitors. *Am. J. Physiol. Heart Circ. Physiol.* **283,** H2029–H2037 (2002).
37. Peng, X., Zhang, C., Alkayed, N. J., Harder, D. R. & Koehler, R. C. Dependency of cortical functional hyperemia to forepaw stimulation on epoxygenase and nitric oxide synthase activities in rats. *J. Cereb. Blood Flow Metab.* **24,** 509–517 (2004).
38. Davis, R. J. *et al.* EP4 prostanoid receptor-mediated vasodilation of human middle cerebral arteries. *Br. J. Pharmacol.* **141,** 580–585 (2004).
39. Takata, F. *et al.* Adrenomedullin-induced relaxation of rat brain pericytes is related to the reduced phosphorylation of myosin light chain through the cAMP/PKA signaling pathway. *Neurosci. Lett.* **449,** 71–75 (2009).
40. Serebryakov, V., Zakharenko, S., Snetkov, V. & Takeda, K. Effects of prostaglandins E₁ and E₂ on cultured smooth muscle cells and strips of rat aorta. *Prostaglandins* **47,** 353–365 (1994).
41. Campbell, W. B., Gebremedhin, D., Pratt, P. F. & Harder, D. R. Identification of epoxyeicosatrienoic acids as endothelium-derived hyperpolarizing factors. *Circ. Res.* **78,** 415–423 (1996).
42. Behm, D. J., Ogbonna, A., Wu, C., Burns-Kurtis, C. L. & Douglas, S. A. Epoxyeicosatrienoic acids function as selective, endogenous antagonists of native thromboxane receptors: identification of a novel mechanism of vasodilation. *J. Pharmacol. Exp. Ther.* **328,** 231–239 (2009).
43. Takano, T. *et al.* Astrocyte mediated control of cerebral blood flow. *Nature Neurosci.* **9,** 260–267 (2006). **Extends, to the *in vivo* situation, the Zonta *et al.* (2003) result that astrocytes control cerebral blood flow.**
44. Mulligan, S. J. & MacVicar, B. A. Calcium transients in astrocyte endfeet cause cerebrovascular constrictions. *Nature* **431,** 195–199 (2004).
45. Blanco, V. M., Stern, J. E. & Filosa, J. Tone-dependent vascular responses to astrocyte-derived signals. *Am. J. Physiol. Heart Circ. Physiol.* **294,** H2855–H2863 (2008).
46. Chuquet, J., Hollender, L. & Nimchinsky, E. A. High-resolution *in vivo* imaging of the neurovascular unit during spreading depression. *J. Neurosci.* **27,** 4036–4044 (2007).
47. Kis, B., Snipes, J. A., Isse, T., Nagy, K. & Busija, D. W. Putative cyclooxygenase-3 expression in rat brain cells. *J. Cereb. Blood Flow Metab.* **23,** 1287–1292 (2003).
48. Hirst, W. D. *et al.* Expression of COX-2 by normal and reactive astrocytes in the adult rat central nervous system. *Mol. Cell. Neurosci.* **13,** 57–68 (1999).
49. Niwa, K., Araki, E., Morham, S. G., Ross, M. E. & Iadecola, C. Cyclooxygenase-2 contributes to functional hyperemia in whisker-barrel cortex. *J. Neurosci.* **20,** 763–770 (2000).
50. Petzold, G. C., Albeanu, D. F., Sato, T. F. & Murthy, V. N. Coupling of neural activity to blood flow in olfactory glomeruli is mediated by astrocytic pathways. *Neuron* **58,** 897–910 (2008).
51. Schummers, J., Yu, H. & Sur, M. Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. *Science* **320,** 1638–1643 (2008).
52. Doengi, M. *et al.* GABA uptake-dependent Ca²⁺ signaling in developing olfactory bulb astrocytes. *Proc. Natl Acad. Sci. USA* **106,** 17570–17575 (2009).
53. Lauritzen, M. Reading vascular changes in brain imaging: is dendritic calcium the key? *Nature Rev. Neurosci.* **6,** 77–85 (2005).
54. Winship, I. R., Plaa, N. & Murphy, T. H. Rapid astrocyte calcium signals correlate

with neuronal activity and onset of the hemodynamic response *in vivo*. *J. Neurosci.* **27**, 6268–6272 (2007).

55. Mathiesen, C., Caesar, K., Akgören, N. & Lauritzen, M. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J. Physiol.* **512**, 555–566 (1998).

56. Enager, P. et al. Pathway-specific variations in neurovascular and neurometabolic coupling in rat primary somatosensory cortex. *J. Cereb. Blood Flow Metab.* **29**, 976–986 (2009).

57. Wang, X. et al. Astrocytic Ca^{2+} signaling evoked by sensory stimulation *in vivo*. *Nature Neurosci.* **9**, 816–823 (2006).

58. Lindauer, U., Megow, D., Schultze, J., Weber, J. R. & Dirnagl, U. Nitric oxide synthase inhibition does not affect somatosensory evoked potentials in the rat. *Neurosci. Lett.* **216**, 207–210 (1996).

59. Yang, G., Chen, G., Ebner, T. J. & Iadecola, C. Nitric oxide is the predominant mediator of cerebellar hyperemia during somatosensory activation in rats. *Am. J. Physiol.* **277**, R1760–R1770 (1999).

60. Caesar, K., Akgören, N., Mathiesen, C. & Lauritzen, M. Modification of activity-dependent increases in cerebellar blood flow by extracellular potassium in anaesthetized rats. *J. Physiol.* **520**, 281–292 (1999).

61. Golanov, E. V. & Reis, D. J. Nitric oxide and prostanoids participate in cerebral vasodilation elicited by electrical stimulation of the rostral ventrolateral medulla. *J. Cereb. Blood Flow Metab.* **14**, 492–502 (1994).

62. Hoffmeyer, H. W., Enager, P., Thomsen, K. J. & Lauritzen, M. J. Nonlinear neurovascular coupling in rat sensory cortex by activation of transcallosal fibers. *J. Cereb. Blood Flow Metab.* **27**, 575–587 (2007).

63. Akgören, N., Mathiesen, C., Rubin, I. & Lauritzen, M. Laminar analysis of activity-dependent increases of CBF in rat cerebellar cortex: dependence on synaptic strength. *Am. J. Physiol.* **273**, H1166–H1176 (1997).

64. Roman, R. J. P-450 metabolites of arachidonic acid in the control of cardiovascular function. *Physiol. Rev.* **82**, 131–185 (2002).

65. Fujimoto, Y., Uno, E. & Sakuma, S. Effect of reactive oxygen and nitrogen species on cyclooygenase-1 and -2 activities. *Prostaglandins Leukot. Essent. Fatty Acids* **71**, 335–340 (2004).

66. Sun, C. W., Falck, J. R., Okamoto, H., Harder, D. R. & Roman, R. J. Role of cGMP versus 20-HETE in the vasodilator response to nitric oxide in rat cerebral arteries. *Am. J. Physiol. Heart Circ. Physiol.* **279**, H339–H350 (2000).

67. Stuehr, D. J., Santolini, J., Wang, Z., Wei, C. & Adak, S. Update on mechanism and catalytic regulation in the NO synthases. *J. Biol. Chem.* **279**, 36167–36170 (2004).

68. Harder, D. R. et al. Identification of a putative microvascular oxygen sensor. *Circ. Res.* **79**, 54–61 (1996).

69. Jūránek, I., Suzuki, H. & Yamamoto, S. Purification, characterization and selective inhibition of human prostaglandin G/H synthase 1 and 2 expressed in the baculovirus system. *Biochim. Biophys. Acta* **1436**, 509–518 (1999).

70. Hall, C. N. & Attwell, D. Assessing the physiological concentration and targets of nitric oxide in brain tissue. *J. Physiol.* **586**, 3597–3615 (2008).

71. Caesar, K. et al. Glutamate receptor-dependent increments in lactate, glucose and oxygen metabolism evoked in rat cerebellum *in vivo*. *J. Physiol.* **586**, 1337–1349 (2008).

72. Hamilton, N. B., Attwell, D. & Hall, C. N. Pericyte-mediated regulation of capillary diameter: a component of neurovascular coupling in health and disease. *Front. Neuroenergetics* **2**, 5 (2010).

73. Shepro, D. & Morel, N. M. Pericyte physiology. *FASEB J.* **7**, 1031–1038 (1993).

74. Puro, D. G. Physiology and pathobiology of the pericyte-containing retinal microvasculature: new developments. *Microcirculation* **14**, 1–10 (2007).

75. Peppiatt, C. M., Howarth, C., Mobbs, P. & Attwell, D. Bidirectional control of CNS capillary diameter by pericytes. *Nature* **443**, 700–704 (2006).

76. Yemişci, M. et al. Pericyte contraction induced by oxidative-nitrative stress impairs capillary reflow despite successful opening of an occluded cerebral artery. *Nature Med.* **15**, 1031–1037 (2009).
    Shows that pericyte constriction decreases blood flow after stroke.

77. Lovick, T. A., Brown, L. A. & Key, B. J. Neurovascular relationships in hippocampal slices: physiological and anatomical studies of mechanisms underlying flow-metabolism coupling in intraparenchymal microvessels. *Neuroscience* **92**, 47–60 (1999).

78. Lu, K. et al. Cerebral autoregulation and gas exchange studied using a human cardiopulmonary model. *Am. J. Physiol. Heart Circ. Physiol.* **286**, H584–H601 (2004).

79. Boas, D. A., Jones, S. R., Devor, A., Huppert, T. J. & Dale, A. M. A vascular anatomical network model of the spatio-temporal response to brain activation. *Neuroimage* **40**, 1116–1129 (2008).

80. Fiser, J., Chiu, C. & Weliky, M. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* **431**, 573–578 (2004).

81. Schölvinck, M., Howarth, C. & Attwell, D. The cortical energy needed for conscious perception. *Neuroimage* **40**, 1460–1468 (2008).

82. Lin, A. L., Fox, P. T., Hardies, J., Duong, T. Q. & Gao, J. H. Nonlinear coupling between cerebral blood flow, oxygen consumption, and ATP production in human visual cortex. *Proc. Natl Acad. Sci. USA* **107**, 8446–8451 (2010).
    Important quantification of the relative magnitudes of stimulus-induced changes in blood flow, O_2 use and ATP generation.

83. Fox, P. T., Raichle, M. E., Mintun, M. A. & Dence, C. Nonoxidative glucose consumption during focal physiologic neural activation. *Science* **241**, 462–464 (1988).

84. Madsen, P. L., Cruz, N. F., Sokoloff, L. & Dienel, G. A. Cerebral oxygen/glucose ratio is low during sensory stimulation and rises above normal during recovery: excess glucose consumption during stimulation is not accounted for by lactate

85. Mangia, S. et al. Metabolic and hemodynamic events after changes in neuronal activity: current hypotheses, theoretical predictions and *in vivo* NMR experimental findings. *J. Cereb. Blood Flow Metab.* **29**, 441–463 (2009).

86. Buxton, R. B. & Frank, L. R. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* **17**, 64–72 (1997).

87. Leithner, C. et al. Pharmacological uncoupling of activation induced increases in CBF and $CMRO_2$. *J. Cereb. Blood Flow Metab.* **30**, 311–322 (2010).

88. Uğurbil, K. et al. Magnetic resonance studies of brain function and neurochemistry. *Annu. Rev. Biomed. Eng.* **2**, 233–260 (2000).

89. Attwell, D. & Iadecola, C. The neural basis of functional brain imaging signals. *Trends Neurosci.* **25**, 621–625 (2002).

90. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 1517–1531 (2001).

91. Markram, H., Lübke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol.* **500**, 409–440 (1997).

92. Hillman, E. M. et al. Depth-resolved optical imaging and microscopy of vascular compartment dynamics during somatosensory stimulation. *Neuroimage* **35**, 89–104 (2007).

93. Hall, C. N. & Garthwaite, J. What is the real physiological NO concentration *in vivo*? *Nitric Oxide* **21**, 92–103 (2009).

94. Lauritzen, M. Pathophysiology of the migraine aura. The spreading depression theory. *Brain* **117**, 199–210 (1994).

95. Fabricius, M. et al. Cortical spreading depression and peri-infarct depolarization in acutely injured human cerebral cortex. *Brain* **129**, 778–790 (2006).

96. Dohmen, C. et al. Spreading depolarizations occur in human ischemic stroke with high incidence. *Ann. Neurol.* **63**, 720–728 (2008).

97. Dreier, J. P. et al. Delayed ischaemic neurological deficits after subarachnoid haemorrhage are associated with clusters of spreading depolarizations. *Brain* **129**, 3224–3237 (2006).

98. Hansen, A. J. & Zeuthen, T. Extracellular ion concentrations during spreading depression and ischemia in the rat brain cortex. *Acta Physiol. Scand.* **113**, 437–445 (1981).

99. Van Harreveld, A. & Kooiman, M. Amino acid release from the cerebral cortex during spreading depression and asphyxiation. *J. Neurochem.* **12**, 431–439 (1965).

100. Barbour, B., Brew, H. & Attwell, D. Electrogenic glutamate uptake in glial cells is activated by intracellular potassium. *Nature* **335**, 433–435 (1988).

101. Piilgaard, H. & Lauritzen, M. Persistent increase in oxygen consumption and impaired neurovascular coupling after spreading depression in rat neocortex. *J. Cereb. Blood Flow Metab.* **29**, 1517–1527 (2009).
     Quantifies changes in energy use, blood flow and neurovascular coupling after spreading depression.

102. Takano, T. et al. Cortical spreading depression causes and coincides with tissue hypoxia. *Nature Neurosci.* **10**, 754–762 (2007).

103. Hashemi, P. et al. Persisting depletion of brain glucose following cortical spreading depression, despite apparent hyperaemia: evidence for risk of an adverse effect of Leão's spreading depression. *J. Cereb. Blood Flow Metab.* **29**, 166–175 (2009).

104. Busija, D. W., Bari, F., Domoki, F., Horiguchi, T. & Shimizu, K. Mechanisms involved in the cerebrovascular dilator effects of cortical spreading depression. *Prog. Neurobiol.* **86**, 379–395 (2008).

105. Fabricius, M., Akgören, N. & Lauritzen, M. Arginine–nitric oxide pathway and cerebrovascular regulation in cortical spreading depression. *Am. J. Physiol.* **269**, H23–H29 (1995).

106. Wahl, M., Schilling, L., Parsons, A. A. & Kaumann, A. Involvement of calcitonin gene-related peptide (CGRP) and nitric oxide (NO) in the pial artery dilatation elicited by cortical spreading depression. *Brain Res.* **637**, 204–210 (1994).

107. Wahl, M., Lauritzen, M. & Schilling, L. Changes of cerebrovascular reactivity after cortical spreading depression in cats and rats. *Brain Res.* **411**, 72–80 (1987).

108. Scheckenbach, K. E., Dreier, J. P., Dirnagl, U. & Lindauer, U. Impaired cerebrovascular reactivity after cortical spreading depression in rats: restoration by nitric oxide or cGMP. *Exp. Neurol.* **202**, 449–455 (2006).

109. Ames, A. III, Wright, R. L., Kowada, M., Thurston, J. M. & Majno, G. Cerebral ischaemia. II. The no-reflow phenomenon. *Am. J. Pathol.* **52**, 437–453 (1968).

110. Nelson, C. W., Wei, E. P., Povlishock, J. T., Kontos, H. A. & Moskowitz, M. A. Oxygen radicals in cerebral ischemia. *Am. J. Physiol.* **263**, H1356–H1362 (1992).

111. Hauck, E. F., Apostel, S., Hoffmann, J. F., Heimann, A. & Kempski, O. Capillary flow and diameter changes during reperfusion after global cerebral ischemia studied by intravital video microscopy. *J. Cereb. Blood Flow Metab.* **24**, 383–391 (2004).

112. Theilen, H., Schröck, H. & Kuschinsky, W. Gross persistence of capillary plasma perfusion after middle cerebral artery occlusion in the rat brain. *J. Cereb. Blood Flow Metab.* **14**, 1055–1061 (1994).

113. Iadecola, C. & Zhang, F. Nitric oxide-dependent and -independent components of cerebrovasodilation elicited by hypercapnia. *Am. J. Physiol.* **266**, R546–R552 (1994).

114. Wagerle, L. C. & Mishra, O. P. Mechanism of CO_2 response in cerebral arteries of the newborn pig: role of phospholipase, cyclooxygenase, and lipoxygenase pathways. *Circ. Res.* **62**, 1019–1026 (1988).

115. Kågström, E., Smith, M. L. & Siesjö, B. K. Cerebral circulatory responses to

hypercapnia and hypoxia in the recovery period following complete and incomplete cerebral ischemia in the rat. *Acta Physiol. Scand.* **118,** 281–291 (1983).

116. Zou, M. H., Leist, M. & Ullrich, V. Selective nitration of prostacyclin synthase and defective vasorelaxation in atherosclerotic bovine coronary arteries. *Am. J. Pathol.* **154,** 1359–1365 (1999).

117. Fleming, I. Cytochrome P450 epoxygenases as EDHF synthase(s). *Pharmacol. Res.* **49,** 525–533 (2004).

118. Sun, J., Druhan, L. J. & Zweier, J. L. Dose dependent effects of reactive oxygen and nitrogen species on the function of neuronal nitric oxide synthase. *Arch. Biochem. Biophys.* **471,** 126–133 (2008).

119. Kaur, J., Zhao, Z., Klein, G. M., Lo, E. H. & Buchan, A. M. The neurotoxicity of tissue plasminogen activator? *J. Cereb. Blood Flow Metab.* **24,** 945–963 (2004).

120. Nicole, O. *et al.* The proteolytic activity of tissue-plasminogen activator enhances NMDA receptor-mediated signaling. *Nature Med.* **7,** 59–64 (2001).

121. Park, L. *et al.* Key role of tissue plasminogen activator in neurovascular coupling. *Proc. Natl Acad. Sci. USA* **105,** 1073–1078 (2008).
   Suggests that tPA, as used clinically to clear clots from blocked vessels, has a role in neurovascular coupling.

122. Armstead, W. M., Cines, D. B. & Al-Roof Higazi, A. Altered NO function contributes to impairment of uPA and tPA cerebrovasodilation after brain injury. *J. Neurotrauma* **21,** 1204–1211 (2004).

123. Cipolla, M. J., Lessov, N., Clark, W. M. & Haley, E. C. Jr. Postischemic attenuation of cerebral artery reactivity is increased in the presence of tissue plasminogen activator. *Stroke* **31,** 940–945 (2000).

124. Johnson, N. A. *et al.* Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: initial experience. *Radiology* **234,** 851–859 (2005).

125. Ruitenberg, A. *et al.* Cerebral hypoperfusion and clinical onset of dementia: the Rotterdam study. *Ann. Neurol.* **57,** 789–794 (2005).

126. Park, L. *et al.* Aβ-induced vascular oxidative stress and attenuation of functional hyperemia in mouse somatosensory cortex. *J. Cereb. Blood Flow Metab.* **24,** 334–342 (2004).

127. Kuchibhotla, K. V., Lattarulo, C. R., Hyman, B. T. & Bacskai, B. J. Synchronous hyperactivity and intercellular calcium waves in astrocytes in Alzheimer mice. *Science* **323,** 1211–1215 (2009).

128. Chow, N. *et al.* Serum response factor and myocardin mediate arterial hypercontractility and cerebral blood flow dysregulation in Alzheimer's phenotype. *Proc. Natl Acad. Sci. USA* **104,** 823–828 (2007).

129. D'Esposito, M., Deouell, L. Y. & Gazzaley, A. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nature Rev. Neurosci.* **4,** 863–872 (2003).

130. Geraldes, P. *et al.* Activation of PKC-δ and SHP-1 by hyperglycemia causes vascular cell apoptosis and diabetic retinopathy. *Nature Med.* **15,** 1298–1306 (2009).

Myelination and support of axonal integrity by glia

Klaus-Armin Nave[1]

The myelination of axons by glial cells was the last major step in the evolution of cells in the vertebrate nervous system, and white-matter tracts are key to the architecture of the mammalian brain. Cell biology and mouse genetics have provided insight into axon–glia signalling and the molecular architecture of the myelin sheath. Glial cells that myelinate axons were found to have a dual role by also supporting the long-term integrity of those axons. This function may be independent of myelin itself. Myelin abnormalities cause a number of neurological diseases, and may also contribute to complex neuropsychiatric disorders.

Glial cells outnumber neurons in the human brain and are involved in almost all neural functions, but for decades they received relatively little scientific attention. This is both a cause and a consequence of the poor understanding of what glial cells do. New technologies available to neurobiologists have now provided unexpected insight into glial-cell function, and have greatly expanded research.

A unique specialization of glia in vertebrates is the deposition of myelin. The ability of oligodendrocytes in the central nervous system (CNS) and Schwann cells in the peripheral nervous system (PNS) to wrap long segments of axons with a multilayered sheath of extended cell membrane (Box 1), and to assemble a complex seal with the axon surface that defines the nodes of Ranvier between long axon segments with myelin (termed 'internodes'), leads to one of the most spectacular and intimate cell–cell interactions in the nervous system (for recent reviews, see refs 1–4).

Myelin was first understood to enable 'saltatory' impulse propagation in axons more than 60 years ago (before it was recognized by electron microscopists to be a specialized outgrowth of glia[5]), and this function is a key concept of neurophysiology. Other functions of glial cells are still not completely understood. We do not even know, at a morphological descriptive level, how myelin is deposited. This will be resolved only with the use of new genetic tools and imaging techniques.

I will begin this Review with a look at myelin evolution and the unique physiology of a myelinated brain, before coming to the unexpected functions of myelinating glia in axon support. This will lead to a discussion of myelin diseases, and the possible role of myelination in higher brain functions. In separate boxes, I will summarize knowledge of the molecular architecture of the axon–myelin unit in the CNS, and the molecular mechanism by which axonal signals control myelination, which is, so far, better understood in the PNS.

From axon–associated glia to myelination

The myelination of axons was the last true 'invention' of vertebrate evolution in the cellular architecture of the nervous system, thought to have occurred in placoderms[6,7] — that is, fish with a hinged jaw. Axons with larger diameters allow more rapid impulse propagation, but it is difficult to increase these diameters when the axons are contained in bone. To overcome this, vertebrates achieved rapid impulse propagation by myelinating small-diameter axons. Improved muscle

control became the basis for the development of complex predatory and escape behaviour, which ultimately drove body size and vertebrate evolution.

Myelin-like ensheathments of axons evolved independently in Annelida (ringed worms), Arthropoda (arthropods) and Chordata (vertebrates and invertebrates with a hollow dorsal neural tube)[8], but they are morphologically distinct. Glial cells that engulf axons without myelinating them are a feature of almost all nervous systems. The physical separation of electrically active axons may limit crosstalk. However, as discussed below, axon-engulfing glia also exert a trophic support that relates to neuronal development, problems associated with axon length, or both.

Myelination follows the basic wiring of the nervous system, and largely occurs postnatally in mammals[9]. Thus, even severe developmental defects of myelin are not lethal to embryos, but instead are a cause of disease.

In the simpler PNS, non-myelinating Schwann cells, or Remak cells, coexist with myelin-forming Schwann cells and engulf multiple small-calibre C-fibre axons (which convey, for example, pain signals) in so-called Remak bundles[10]. Why do we have Remak cells that may represent an ancestral type of axon-associated glia? The functional perturbation of Remak cells in transgenic mice causes progressive C-fibre degeneration and sensory neuropathy[11], suggesting that the cells are required to maintain the integrity of axons. This is presumably a function of all axon-associated glial cells[12].

Oligodendrocytes — which myelinate several axons simultaneously — and Schwann cells — which restrict myelination to one axonal segment — solve similar tasks with an overlapping, but not identical, set of genes. This is a result of some 300 million years of parallel evolution, which is thought to have begun with a glial cell similar to those in unmyelinated fish, such as lamprey. Although the mechanism that led to the evolution of myelin in such jawless fish may never be known, it is intriguing that overexpression of a single axonal growth factor turns non-myelinating into myelinating Schwann cells[13].

The CNS contains no equivalent to Remak cells. Oligodendrocytes arise from a large population of oligodendrocyte precursor cells, which are morphologically complex, unlike stem cells (Box 1). They can be defined by expression of the proteoglycan protein NG2, and have been considered a 'fourth' class of glia in the brain[14]. In retrospect, it is astounding that it took neurobiologists so long to recognize the abundance of these oligodendrocyte lineage cells. The finding that NG2-

[1]Department of Neurogenetics, Max Planck Institute of Experimental Medicine, Hermann-Rein-Strasse 3, 37075 Göttingen, Germany.

BOX 1

Oligodendrocytes and the axon myelin unit

Compact myelin
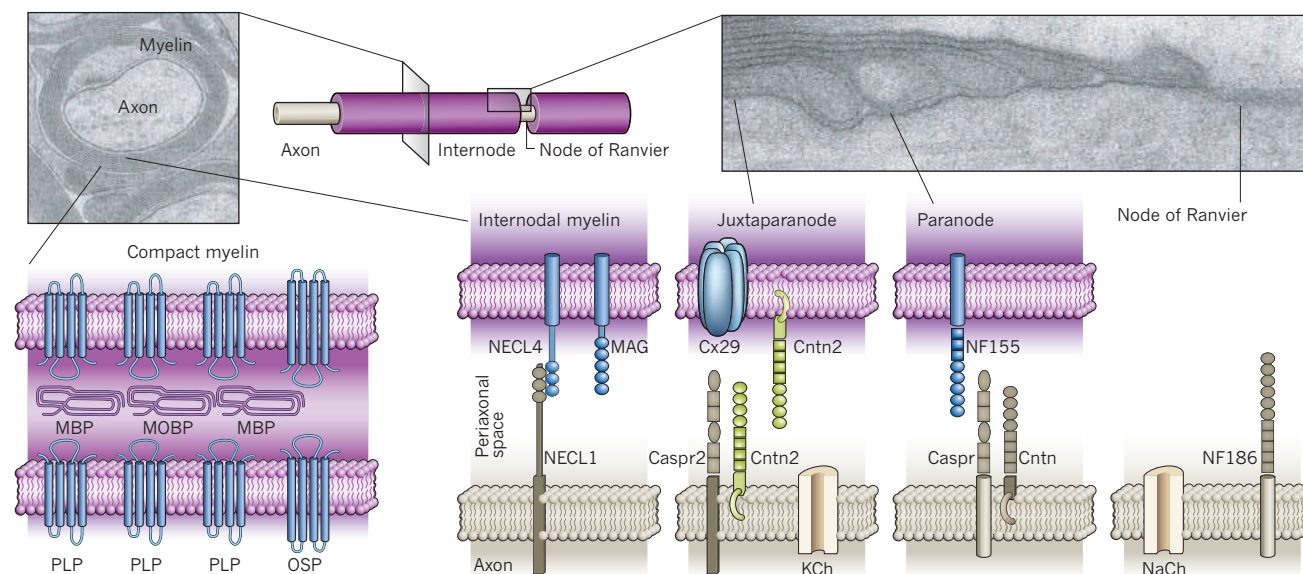
In the rodent central nervous system (CNS), myelin is formed by oligodendrocytes that are derived from morphologically complex precursor cells (oligodendrocyte precursor cells or NG2 cells); the function of these is not well understood[14,83]. Successive stages of oligodendrocyte precursor cell (OPC) differentiation can be visualized *in situ* following the injection of lucifer-yellow dye (LY; see bottom image, adapted from ref. 84).

Mature oligodendrocytes make myelin by wrapping axons with their own cell membrane in a spiral shape, which eventually becomes a multilayered sheath covering a long segment of axon. Although oligodendrocytes are morphologically distinct from the Schwann cells in the peripheral nervous system, axon–myelin units are similar and have been illustrated in the top image as a 'canonical' myelin sheath. Compact myelin is shown by electron microscopy to have a periodic ultrastructure (the inset shows a myelinated optic-nerve axon in cross section). The compaction requires the abundant expression of structural proteins, such as proteolipid protein (PLP) and myelin basic protein (MBP), in the CNS. The functions of most myelin-associated proteins, which have only recently been identified by proteomics[85], are not well understood. A fraction of these proteins may reflect intracellular biogenesis and transport of myelin components that have a slow turnover rate. Myelin membranes are very rich in lipids. Cholesterol, specifically, is rate-limiting for myelin biogenesis[86].

The myelinated axon segments are flanked by nodes of Ranvier, the sites at which Na$^+$ channels are concentrated and action potentials generated. Here, a specialized axon–glia contact zone (paranodal junction) also seals the internodal periaxonal space from the outside milieu. The molecular architecture of the zone is reminiscent of synaptic junctions, and builds on a battery of glia and axonal adhesion proteins that are linked by axonal scaffolding proteins (not shown) to Na$^+$ channels and K$^+$ channels (in the juxtaparanodal region). For details, see ref. 4.

Caspr, contactin-associated protein; Cntn, contactin (Cntn2 is also known as Tag1); Cx29, connexin 29 kDa; KCh, fast potassium channels; MAG, myelin-associated glycoprotein; MBP, myelin basic protein; MOBP, myelin oligodendrocyte basic protein; NaCh, voltage-gated sodium channels; NECL, nectin-like protein/synCAM; NF155/186, neurofascin 155 kDa/186 kDa; OSP, oligodendrocyte-specific protein; PLP, proteolipid protein.

OPC / NG2 cell

Pre-oligodendrocyte

Mature oligodendrocyte
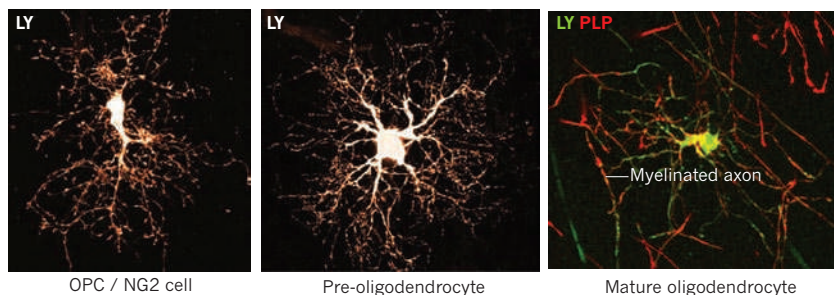
expressing cells also express glutamate receptors and receive transient synaptic input from unmyelinated axons[15] fits earlier observations that neuronal activity modulates myelination. One laboratory has even reported that a subset of NG2 cells can fire action potentials[16], blurring the distinction between neurons and glia.

The benefit of myelinated axons

In the PNS, where neuron–glia interactions are simpler, neurons control glia remotely for their own advantage. The survival, proliferation and terminal differentiation of Schwann cells, Remak cells and their common precursors are largely controlled by one axonal growth factor,

termed neuregulin-1 type III (ref. 17). In the CNS, the proliferation and differentiation of oligodendrocyte lineage cells is controlled by further growth factors and cytokines, such as platelet-derived growth factor, brain-derived neurotrophic factor, ciliary neurotrophic factor (CNTF) and leukaemia inhibitory factor. However, not all of these factors are neuronal, and there is no known axonal signal that drives myelination of the very axon that presents it (Box 2).

Neurons benefit from the myelination of their axons in two ways. First, in electrically active but unmyelinated fibres, the restoration of ion gradients by the Na$^+$, K$^+$-ATPase consumes a large fraction of available ATP. Myelination strongly reduces this energy consumption,

because action potentials and ion currents are restricted to less than 0.5% of the axon's surface. Second, the up to 100-fold increase in conduction velocity (the actual factor depends on the axon diameter) for myelinated neurons allowed complex yet compact higher nervous systems to evolve.

Increasing the resistance (and lowering the capacitance) of axonal membranes is not the only function of oligodendrocytes and Schwann cells. There is growing awareness that myelinating glia communicate lifelong with axons, and that glia are required for the long-term integrity and survival of axons[18–22]. This vital function, which may have a neurotrophic component, is independent of myelin itself and relevant to a broad spectrum of human myelin diseases (see 'Myelin diseases leading to axonal degeneration').

Axon–glia interactions beyond signalling

To understand axon–glia interactions, one must consider cell size. Neurons are extremists, as is shown by a scaled up model. The 1-μm axon of a cortical neuron, projecting 100 cm into the corticospinal tract, can be compared to a 4-m-wide subway tube that is 4,000 km long. To extend the metaphor, fast axonal transport (of mitochondria, at 5 μm s^{-1}) is analogous to a train travelling at 70 km h^{-1}, whereas slow axonal transport (of proteins, at 2 mm day^{-1}) is, by analogy, slower than a pedestrian (0.5 km h^{-1}). Neurons with long axons, therefore, present a logistical problem, and are a vulnerable bottleneck for the entire nervous system. Axon dimensions may even limit the sizes of the largest vertebrates[23]. Neurological diseases that are known to have a primary origin in the glia cause length-dependent axon loss (see 'Myelin diseases leading to axonal degeneration'), showing that long axons profit from the intimate interaction with local glial cells[24].

Axons and the ensheathing glia interact bidirectionally and throughout life. The finding that PNS axons in dysmyelinated trembler mice, which have a mutation in the peripheral myelin protein 22 (*Pmp22*) gene, are reduced in calibre[25] was early evidence for glia-to-axon (glia–axon) signalling. Similarly, dysmyelinated axons in the CNS of shiverer mice (which lack expression of myelin basic protein, MBP) are thinner than in wild-type mice[26]. Hypophosphorylated neurofilaments suggest that abnormal activity of kinases and/or phosphatases underneath immature myelin further tightens the axonal bottleneck.

The glial signals that modulate axon size are not well defined. One candidate is myelin-associated glycoprotein (MAG), a non-compact myelin protein and a member of the immunoglobulin superfamily[27]. The protein can be shed by proteolytic cleavage, and has been studied as an inhibitor of axonal regeneration. MAG-deficient mice are fully myelinated but show decreased axon calibres and neurofilament spacing, which are probable causes of progressive axonal loss[22]. MAG-mutant mice are more sensitive to the poorly understood neurotoxin acrylamide than are the wild type, and soluble recombinant MAG reduces axon damage in cellular models[28]. The axonal receptor for this protective function is unrelated to the neurite outgrowth inhibitor A (NogoA) receptor complex, which mediates growth-cone collapse.

Cytokines derived from Schwann cells, such as CNTF[29] and erythropoietin[30], also enhance axon survival, at least in experiments of nerve regeneration. Thus, glial signalling molecules provide a line of endogenous neuroprotection. In early development, signalling between neurons and Schwann cells is even more vital. Death of the Schwann cells in embryos, caused by disrupted expression of glial neuregulin-1 type III receptors (ErbB3)[31], leads to secondary loss of motor neurons and most dorsal-root ganglia neurons, resulting in perinatal death in mice.

Axonal integrity requires more than glia signalling molecules. Myelin assembly seems to be normal in *Plp1*-null mice, which lack expression of proteolipid protein (PLP), an abundant tetraspan membrane protein in CNS myelin, whereas natural *Plp1* mutants, such as jimpy and rumpshaker mice, exhibit dysmyelination caused by the toxicity of a misfolded protein[32]. Except for minor ultrastructural abnormalities of myelin and reduced physical stability[33], the

PLP-null mutants develop normally and are long-lived. In contrast to dysmyelinated shiverer mutants, the myelinated axons even exhibit normal neurofilament spacing and are developmentally mature. At 3 months of age, however, axons in almost all white-matter tracts in PLP mutants develop swellings ('spheroids') filled with membranous organelles and phosphorylated neurofilaments[18] (similar to those in Fig. 1a, b). As swellings enlarge, the myelin sheath often retracts and is lost over swollen regions. Later, many affected axons in *Plp1*-null mice undergo distal Wallerian degeneration. Within a few months, mice develop a progressive ataxia and die prematurely, without obvious myelin loss but with signs of severe neurodegeneration[18]. Small-calibre axons, which may have the lowest intra-axonal energy reserves, are predominantly affected. Before axonal degeneration there is perturbation of fast axonal transport[19], with possible consequences for brain function as discussed in 'Myelin plasticity and higher cognitive functions'.

It is intriguing that shiverer mice exhibit no axonal degeneration, although they are severely dysmyelinated[18]. More importantly, axonal swellings recur in *Mbp Plp1* double mutants, an observation that points to a myelin-independent role of PLP in axon support. The first functional insight into the mechanism came with an observation that other myelin proteins are less abundant in, or absent from, PLP-deficient myelin[34]. This lack is a failure of post-translational transport into the myelin compartment. The near-absence of nicotinamide adenine dinucleotide-dependent deacetylase sirtuin-2 (SIRT2), a protein expressed by oligodendrocytes and Schwann cells[34–36], is most striking. Although the physiological targets of this enzyme are not known, the phenotype of SIRT2-mutant mice suggests that it has a vital function in neuroprotection (B. Kasapoglu, H. Werner, L. Guarente and K.-A.N., unpublished observations). Recent reports that acetylation regulates the key steps of intermediate metabolism[37] are compatible with SIRT2 having a role in adapting the metabolism of glia.

Specific genetic targeting can be expected to deliver direct insight into the role of the oligodendroglial intermediate metabolism. For example, peroxisomes, which are abundant in oligodendrocytes, are involved in fatty-acid β-oxidation and are crucial for axon function. Conditional mutants have been created, in which oligodendrocytes lack peroxisomal targeting signal 1 receptor (PEX5)-mediated protein import into peroxisomes[38]. During development, these mice assemble myelin that seems normal. However, at about 2 months of age, they exhibit signs of premature neurodegeneration in white-matter tracts and die within one year. Notably, this primary defect of oligodendrocytes causes a secondary neurodegeneration that exceeds demyelination pathology, and is most advanced in the electrically most active subcortical-fibre tracts of the forebrain[38] (Fig. 1c).

More than one gene is responsible for the support of axons by myelinating glia. For example, the oligodendrocyte-specific protein 2',3'-cyclic nucleotide phosphodiesterase (CNP), which resides in non-compact myelin[39], is thought to bind to RNA and tubulin, and to contribute to oligodendroglial process dynamics *in vitro*[39,40]. Like PLP, CNP is not essential for myelination, but is required for axonal integrity[20] (Fig. 1a, b). *Cnp1*-mutant mice are more severely affected than are *Plp1*-null mice, with the first degenerating axons detected after a just few weeks[21]. Moreover, swellings of the inner tongue emerge very early in life[21]. These point to the perturbation of transport processes within the myelin sheath that could also depend on intact tubular tracts.

I have argued that myelination itself necessitates the support of glia, because for long axons the physical insulation decreases rapid and unlimited metabolic exchange with the extracellular milieu[24]. However, non-myelinating Schwann cells also protect the survival of sensory axons, in line with the hypothesis that axonal support is myelin-independent. Proof of this principle is the phenotype of transgenic mice that express truncated ErbB4 receptors in Remak cells[11], which develop a progressive sensory neuropathy with the loss of numerous unmyelinated C-fibre axons.

BOX 2

Axon–glia signalling and phosphatidylinositol-(3,4,5)-trisphosphate as a nodal point of myelin growth

In the peripheral nervous system, myelination is under the strict control of axons. Among several growth factors, membrane-bound neuregulin-1 type III (NRG1-III) regulates almost all steps of Schwann-cell differentiation[17], including the control of myelin-sheath thickness[87], which is a function of axon size. It is thought that signalling between NRG1-III and the epidermal growth factor receptor (ErbB) family of proteins occurs at the glia–axon interface and activates several second-messenger cascades, including phosphatidylinositol-3-kinase (PI3K) and mitogen-activated protein kinase pathways, focal adhesion kinase, and Ca²⁺ (reviewed in ref. 88). Depicted here is a schematic magnification of paranodal loops with glial ErbB receptors and the sequential activation of PI3K–serine-threonine-specific protein kinase AKT/PKB (AKT1)–mammalian target of rapacin (mTOR) kinases, leading to the activation of myelin-associated genes in the Schwann-cell nucleus. Extracellular matrix components and converging integrin-mediated signalling[88] are

not shown. Lipid synthesis is the primary transcriptional response following NRG1-III stimulation in vitro[89] and in vivo (M. Schwab and K.-A.N., unpublished observations).

The formation of phosphatidylinositol-(3,4,5)-trisphosphate (PIP₃) by PI3K polarizes Schwann cells and stimulates cellular RhoGTPases, actin dynamics and process outgrowth, all of which might be required for membrane protrusions and spiral wrapping of axons[90]. The formation of PIP₃ as a lipid second messenger and nodal point of myelination is tightly controlled and antagonized by the phosphatase and tensin homologue (PTEN) protein. Associated with a scaffolding protein (mammalian discs large homolog 1, DLG1), this lipid phosphatase acts as a brake on myelin growth, preventing hypermyelination and pathological outfoldings[90,91]. In adult nerves, Schwann cells require axons to express the prion protein to prevent demyelination and neuropathy[92].

Oligodendrocytes are different to Schwann cells; they differentiate well in the absence of axonal signals (at least in vitro). The effect of neurotrophins on myelination by oligodendrocytes is also distinct from the effect on Schwann cells[93]. In vivo, NRG1 and ErbB signalling is largely dispensable for myelination in the central nervous system (CNS)[94], prompting a search for alternative axonal signals. Insulin-like growth factor 1 (IGF-1) and astroglial leukaemia inhibitory factor (LIF) stimulate CNS myelination[95,96], but are not associated with specific axons. AKT and mTOR drive hypermyelination in both Schwann cells and oligodendrocytes[90,97], and transgenic NRG1 overexpression in CNS neurons stimulates some hypermyelination[94]; these observations support a signalling system similar to NRG1. On the other hand, oligodendrocytes in culture can wrap even chemically fixed axons[98]. Thus, neither complex instructive signals nor electrical activity seems to be essential for myelination. This leaves room for the idea that oligodendrocytes are stimulated by a plethora of unconfined growth factors to myelinate by default, but are locally restricted by inhibitory cues, such as polysialylated neural cell adhesion molecule (NCAM)[99].

BACE, β-site amyloid precursor protein-cleaving enzyme 1; EGR2, early growth response protein 2 (also known as Krox20); ErbB2/ErbB3, heterodimeric NRG1 receptor tyrosine kinases; IGF-1R, IGF-1 receptor; PIP₂, phosphatidylinositol-4,5-bisphosphate; SREBP, sterol regulatory element binding protein. Dashed lines indicate pathways that are not well-established.
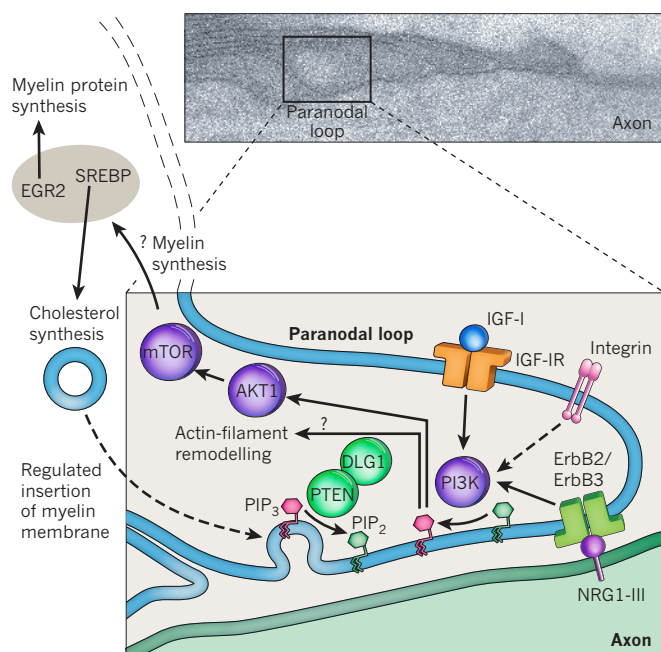
Some myelin defects trigger inflammation

Mild overexpression of wild-type *Plp1* in transgenic mice causes a combination of dysmyelination and demyelination in the CNS, leading to premature death[41,42]. The disease mechanism exceeds endoplasmic-reticulum stress and includes lipid-trafficking abnormalities[43]. PLP overexpression models a frequent subform of the human *PLP1* disease (see Pelizaeus–Merzbacher disease below) and causes late-onset axonal degeneration[44]. The mechanism of axon loss is unclear, but a secondary inflammation (Fig. 1d), including microgliosis and T-cell infiltration, clearly contributes to the severity of the disease[45]. Invading CD8⁺ T cells are not just bystanders, because on a recombinant activating gene-1 mutant background (that is, in the absence of functional B and T cells) a significant amelioration of pathology is seen. *Plp1* transgenic mice with T cells lacking the proteins perforin or granzyme B also fail to show the full-blown phenotype[46]. In the optic nerve of *Plp1* transgenics, it is striking that axonal segments at the already demyelinated (retinal) end are less visibly affected by swellings than are the distal axonal segments,

which are associated with residual myelin and inflammation[47]. Thus, with respect to axon function, 'no myelin' may be better than 'bad myelin', with inflammation posing an extra burden for axonal metabolism.

Although T-cell infiltration is a feature of other neurodegenerative conditions, oligodendrocyte dysfunction may be of specific relevance for inflammation. The strongest secondary immune response was observed in the white matter of conditional *Pex5*-mutant mice (Fig. 1c), including activated and clonally expanded CD8⁺ T cells and perivascular B cells (ref. 38 and B. Barrette, C. Kassmann and K.-A.N., unpublished data). Myelin-lipid turnover creates arachidonic acid and eicasonoids, which include lymphocyte chemoattractants that are degraded in peroxisomes[48]. Thus, abnormal myelin turnover and the failure to degrade inflammatory mediators may contribute to neuroinflammation.

In these myelin mutants, microgliosis and the recruitment of CD8⁺ T cells by genetically perturbed (but live) oligodendrocytes may trigger a vicious cycle of degeneration and inflammation. This is a working hypothesis relevant to human neurological disease.

Myelin diseases leading to axonal degeneration

The differentiation of myelinating glia largely occurs postnatally. Thus, even major genetic defects result in a developmental disorder rather than death of the embryo. The clinically most important disorders are acquired demyelinating diseases, which present during adolescence and in adult life.

Multiple sclerosis

Multiple sclerosis (MS) is the prototype of immune-mediated demyelination[49], but its trigger and the primary cause of autoimmunity remain unknown. Lymphocytes and monocytes cross the blood–brain barrier and invade the brain. Focal areas of inflammation and immune attacks destroy oligodendrocytes and myelin sheaths, but rarely dissect axons directly. Instead, they lead to lesions in white matter, detectable by magnetic resonance imaging (MRI), that turn into oedematous plaques. There is also major loss of myelinated fibres in cortical grey matter; this loss has been underestimated, because it does not have the same degree of cellular infiltration[50]. Only a fraction of MS lesions become clinically apparent. Often, the acute signs result from axonal conduction blocks, but are reversible as inflammation and oedema resolve, sodium channels reorganize and remyelination occurs. After years of 'relapsing–remitting' MS, most patients enter the secondary phase: 'progressive' MS, characterized by increasing disability without acute attacks or remissions[49]. Although immune modulators, such as β-interferon, are given to reduce the rate of relapse, there is no similar therapy against secondary, progressive MS. It is thought that, at this stage of disease, axon degeneration causes persistent disability[51]. Although old lesions are never free of inflammation, the correlation with the swelling and loss of axons becomes weaker. MS is probably heterogeneous and triggered in susceptible individuals by the late consequence of an early infection that turns into a vicious cycle of inflammation and neurodegeneration. In some MS patients, progressive neurodegeneration may be a distinct disease mechanism[51].

Research into the mechanism of axon loss as the final common pathway in MS is hampered by the absence of true animal models. In mice with experimental allergic encephamyelitis, disease progression is driven by autoreactive CD4[+] T cells, which model only one aspect of neuroinflammation. Also in the genetic models, the infiltration of activated CD8[+] T cells is insufficient to trigger MS-like lesions. It is likely that a combination of inflammation and demyelination creates a high risk of axon loss; the lack of oligodendroglial support occurs when axons require that support the most. Demyelination and new axonal Na[+] channels reduce the axonal energy balance. Inflammation generates reactive oxygen species — such as nitric oxide — and glutamate-mediated excitotoxicity, which further perturbs residual mitochondrial functions. Once ATP is so low that the axonal membrane cannot be repolarized, a 'threshold' is reached that leads to abnormal calcium entry, proteolysis and axonal destruction[52].

Leukodystrophies

Leukodystrophies in the CNS comprise a heterogeneous group of rare inherited disorders of dysmyelination and demyelination, in which oligodendrocytes fail to assemble or to maintain myelin, respectively. Children with perturbed myelination experience a severe delay in reaching basic milestones of motor–sensory and cognitive development. People with leukodystrophies who survive into adolescence develop a secondary degenerative course of disease, which points to irreversible axon loss. The underlying mechanisms are currently being studied in mouse models, and the search concentrates on the possible role of abnormal glia–axon signalling, toxic insults and perturbed metabolic support.

Disease classifications, originally based on histopathology, biochemical markers and MRI, have shifted to genetic criteria. A large proportion of human leukodystrophy genes have now been identified. However, each of the known leukodystrophies covers a broad clinical spectrum, in which disease severity is determined by the specific mutation. Thus,
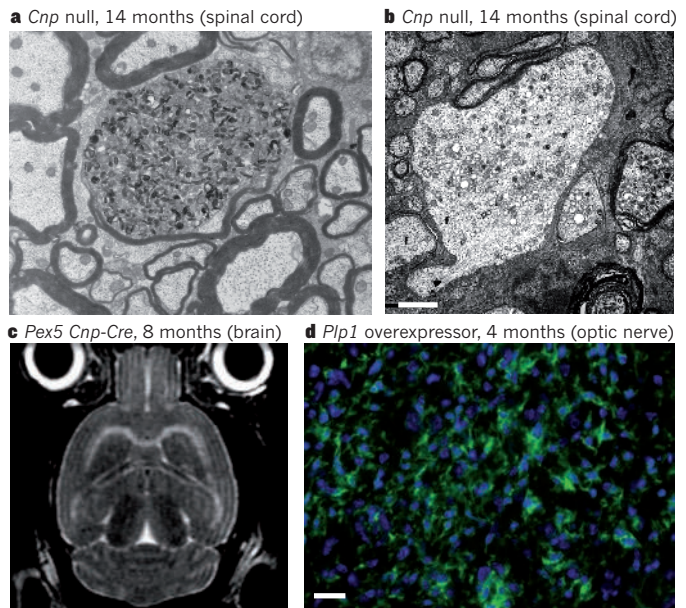
Figure 1 | Oligodendrocyte defects causing axonal degeneration in the central nervous system. a–d, Electron microscopy (**a**, **b**), magnetic resonance imaging (MRI; **c**) and anti-CD45 immunostaining (**d**) of mouse mutants, defined by abnormal expression and (conditional) mutations of oligodendroglial genes. *Cnp*, 2′,3′-cyclic nucleotide 3′ phosphodiesterase; *Cnp-Cre*, Cre-recombinase glial driver line; *Pex5*, peroxisomal biogenesis factor 5; *Plp1*, proteolipid protein 1. Although myelination appears relatively normal in these mutants, axonal transport — mostly of small-calibre axons — is perturbed, leading to Wallerian degeneration and premature death. Swellings contain vesicular cargo and organelles (**a**) and/or neurofilament aggregates (**b**). In more severely affected models[20,38,47], neurodegeneration is associated with monocyte and lymphocyte infiltration (**d**). One model, lacking oligodendroglial peroxisomes, eventually loses most of the subcortical white matter, as visualized by MRI hyperintense signals (**c**). Images are taken from (or related to) refs 20 (**a**), 82 (**b**), 38 (**c**) and 47 (**d**). Ages are indicated. Scale bars, 2 μm (**b**) and 20 μm (**d**).

defects in the same gene can result in early-infantile or late-adult onset, owing to complex combinations of gain- and loss-of-function effects, in combination with modifier genes and epigenetic differences. The dual role of oligodendrocytes in myelination and axonal support becomes most obvious in the late-onset forms.

A prototype of the hypomyelinating leukodystrophies (HLD) is Pelizaeus–Merzbacher disease (PMD or HLD1, Online Mendelian Inheritance in Man (OMIM) accession 312080), which is caused by mutations or duplications of the X chromosome-linked *PLP1* gene[53]. Most mutations cause a misfolding of PLP. The degree to which the polytopic membrane protein escapes endoplasmic-reticulum retention determines the extent of an unfolded protein response, apoptotic oligodendrocyte death and CNS dysmyelination. Clinically, *PLP1*-related disorders span a broad spectrum of disease expression, from early lethality in the connatal forms to the mildest course of disease, pure spastic paraplegia (SPG2). The latter is clinically distinct (OMIM 312920), with relatively few myelin abnormalities. These different phenotypes of *PLP1* diseases have been successfully modelled in natural *Plp1* point mutations, *Plp1* transgenic mice and *Plp1*-null mutants[54].

A 'PMD-like' HLD (HLD2, OMIM 608804)[55] has been defined by mutations in a gap junction protein gene, which encodes the protein connexin 47 (Cx47). This developmental disorder demonstrates a requirement of efficient coupling by gap junctions between oligodendrocytes (with Cx47, Cx32) and astrocytes (Cx43, Cx30), possibly for the transport of metabolites during myelinogenesis[56]. The adult forms (SPG44, OMIM 613206) also demonstrate that pan-glia coupling has an important role in axonal preservation[57].

That astrocytes are important for myelination and for preventing demyelination is demonstrated by Alexander's disease[58] (AXD; OMIM 203450) and mutations of the glial fibrillary acidic protein gene, which encodes an astrocyte-specific intermediate filament. The onset of AXD is variable, and detailed disease mechanisms are unknown. The structural integrity of astrocytes, which provide a link between oligodendrocytes and the blood–brain barrier, seems to be critical for myelin preservation and axonal support.

Leukodystrophy genes do not encode only the proteins specific to glial-cell function. For example, 'vanishing white matter' disease (VWM; OMIM 603896) can be attributed to mutations affecting eukaryotic translation initiation factor (EIF2B) subunits, which comprise a housekeeping protein and translational-initiation factor[59]. Mouse models can be used to explore whether it is the role of EIF2B in the efficient generation of astrocytes[60] or the inability of oligodendrocytes to fine-tune membrane-protein translation (thereby causing endoplasmic-reticulum stress) that explains the clinical and pathological similarities of VWM to AXD.

Some leukodystrophies are degenerative disorders that manifest later in life, affecting white matter that has developed normally. For example, in X-linked adrenoleukodystrophy (X-ALD; OMIM 300100), loss of a peroxisomal transporter protein (ABCD1) interferes with the degradation of very-long-chain fatty acids and causes an inflammatory demyelination[61]. Although the gene is widely expressed, only oligodendrocytes and adrenal cells are affected; this may reflect their unusual degree of lipid metabolism. *Abcd1*-mutant mice fail to show the X-ALD phenotype within a normal lifespan. However, targeting peroxisomal import in myelinating glia[38] creates a very good phenocopy (Fig. 1c). This suggests that inflammatory demyelination and axon loss might be caused by a secondary peroxisomal defect and by organelle dysfunctions, which accumulate over time and only in *ABCD1* mutant oligodendrocytes.

Demyelinating neuropathies

As in the CNS, genetic defects in making or maintaining myelin in the PNS comprise a heterogeneous group of diseases. These are collectively referred to as 'demyelinating' Charcot–Marie–Tooth (CMT) neuropathies, and have been reviewed elsewhere[62,63]. Genetically, there is surprisingly little overlap with leukodystrophies, but the abnormal expression of membrane proteins, such as myelin protein zero (MPZ) and PMP22, in Schwann cells leads to endoplasmic-reticulum stress, followed by dysmyelination and demyelination. Hypertrophic myelin growth is a unique feature of CMT disease. Here, tomacula and myelin outfoldings originate at Schmidt–Lantermann incisures and at para-nodal loops, the latter disrupting the sites of axon–glia interactions. Most neuropathies have been modelled in mutant or transgenic mice.

The true absence of myelin is rare; more frequent are moderate dysmyelinating and demyelinating conditions in which the decrease of nerve-conduction velocity is clinically silent. All relevant symptoms, such as distally pronounced muscle weakness and sensory loss, are attributable to progressive and length-dependent degeneration of axons[62,63]. The role of Schwann cells in axonal integrity, as with that of oligodendrocytes, can be uncoupled from myelination. Rare point mutations in *MPZ* that cause a late-onset axonal form of CMT disease (OMIM 607736), which does not affect conduction velocity and myelination but does cause sensory defects including hearing loss, are proof of this principle[64]. What makes these *MPZ* mutations different from those in patients with the demyelinating form of CMT is not known.

Myelin plasticity and higher cognitive functions

Because brain evolution entailed an increase in the amount of subcortical white matter, one would assume that myelination is necessary for higher cognitive functions. However, except for case reports of psychosis in MS and adult-onset leukodystrophy, myelin diseases
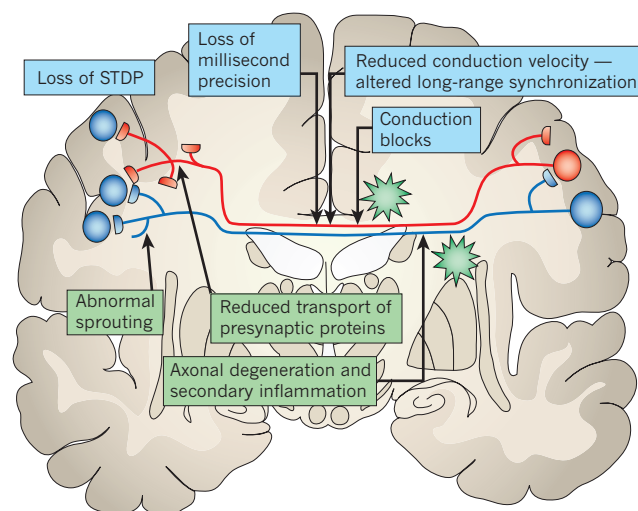
Figure 2 | Oligodendrocyte defects may lead to cognitive impairment. A hypothetical model pointing to possible consequences of reduced myelination, altered axonal diameters, and/or fast axonal transport, when caused by oligodendroglial dysfunction (blue boxes). In this schema, transcallosal cortical projections are myelinated and can establish and maintain long-range oscillations between cortical subfields. Spike-timing-dependent plasticity (STDP) requires millisecond precision, which is myelin dependent. Perturbed fast axonal transport may alter the protein composition of presynaptic terminals. More severe oligodendroglial perturbations, as caused by inflammations (green star-like cells) and demyelination, lead to axonal defects (green boxes), such as conduction blocks and/or irreversible axon loss that clearly disrupt long-range connectivity. Poor intracortical myelination might also trigger abnormal axon sprouting and/or interfere with memory consolidation.

have not been studied well with respect to psychiatric phenotypes. An unbiased search for transcriptional changes in post-mortem brains of schizophrenics has revealed a significant decrease in abundance of RNAs for oligodendrocyte-specific proteins[65]. Although these changes affected the intracortical oligodendrocytes, MRI also revealed the loss of subcortical white matter in patients with schizophrenia, depression and bipolar disorder[66]. Unfortunately, MRI lacks the spatial resolution to distinguish between changes in axon number and myelin thickness. The mechanisms by which oligodendrocytes might contribute to psychiatric disorders such as schizophrenia are not known, and are probably complex (Fig. 2).

There are several reasons why myelination ought to be required for higher brain function. Myelination of cortical association fibres in humans continues into the third decade of life, paralleling cognitive maturation[9]. Even in adults, the acquisition of fine motor skills, such as professional piano playing, has been associated with substantial changes in the white matter in corresponding areas of the motor cortex[67].

Myelination itself shows some plasticity, as was first recognized in electric fish that change myelin-sheath thickness to control the velocity of the axons that trigger electrical discharges (these nerves have differing lengths, but their input to electrocytes must be synchronized with millisecond precision). Adaptations of conduction velocity to synchronize neural activity are also found in mammals, but whether these involve alterations of myelin is unclear. Earlier work has shown that Schwann-cell differentiation is regulated in response to the spiking frequency of associated axons[68], and works through an activity-dependent release of ATP from axons that is sensed by P2 receptors. Similarly, electrical activity of the optic nerve increases the number of oligodendrocyte precursors present in the nerve[69], possibly in response to the release of adenosin[70].

Oligodendrocytes stimulate the regional growth of axonal diameter independently of myelination[71]. This is intriguing, because axon calibre is a major variable in conduction velocity. Because the

electrical activity of axons is detected by glia[70], it is worth exploring whether back-signalling from glia to axons can modulate axon size, for example by neurofilament phosphorylation[72]. Such an adaptation of conduction velocity would be faster and more dynamic than slow changes in myelin-sheath thickness. But how can oligodendrocytes perceive the electrical activity of the axons that they ensheath? In addition to the study of adenosin and P2 receptors[70], much recent work has been devoted to the role of AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and NMDA (N-methyl-D-aspartate) type glutamate receptors on NG2 cells and myelinating oligodendrocytes[73–75]. These could enable the oligodendrocytes to sense the release of glutamate, possibly as a function of spiking activity. The downside to glutamate responsiveness could be a sensitivity of mature oligodendrocytes to excitotoxic injury[76].

Focal myelin loss in the adult brain is followed by remyelination; this plasticity may reflect a physiological function in adult brains. Myelin could stabilize the wiring of axonal terminals — a property that is associated with functional plasticity. This property has been demonstrated in mice lacking the neuronal receptor (NGR) for Nogo, MAG and oligodendrocyte myelin glycoprotein; that is, myelin proteins that inhibit axonal sprouting and regeneration. NGR mutants exhibit a longer critical period, in which ocular dominance columns can be reshaped in the visual cortex[77]. Thus, myelination consolidates specific neural circuitry of the cortex. Indirect evidence suggests that abnormal Nogo and NGR signalling is a risk factor for schizophrenia[78].

Hypomyelination and reduced conduction velocity in long axons limit the distance over which cortical neurons can fire in synchrony[23]. The absence of long-range γ-oscillations between cortical subfields could cause attention deficits[79]. Myelin defects in long axons would also perturb spike-timing-dependent plasticity (STDP), which underlies synaptic strengthening or weakening (long-term potentiation or long-term depression) and possibly long-range circuit refinements[80]. For example, unmyelinated axons that have transcallosal conduction times of about 0.5 s will not maintain the temporal precision in the 10–20 ms range[23] that is critical for STDP.

Degeneration of cortical axons in the presence of apparently normal myelin — for example, that assembled by oligodendrocytes lacking PLP or CNP — has obvious consequences for brain function. The axonal degeneration is preceded by significant perturbation of fast axonal transport[19]. This raises the intriguing possibility that oligodendroglial support is required for the timely delivery of presynaptic components, such as nuclear-encoded enzymes and proteins that control the fast calcium-regulated neurotransmitter release. For example, a 50% gene dosage of Munc18 (mammalian uncoordinated 18 or syntaxin binding protein 1), which encodes a mediator of synaptic vesicle docking, is sufficient to reduce the readily releasable pool size in various synapses[81] and to cause hyperactivity, a behaviour related to schizophrenia in mice. Little is known about how the steady state of this or other presynaptic proteins is controlled in the terminals of long axons, but the effect of reduced transport rates is probably equivalent to that of reduced expression rates.

Conclusion

As in all facets of neurobiology, the myelin field has made unexpected progress in the past few decades, following several breakthroughs in molecular cell biology and mouse genetics. Research on myelin has always been interdisciplinary, with a strong focus on neurological diseases, and is likely to include more psychiatric disorders in the future. Having reached a better understanding of glial-cell development and myelin diseases, fascinating new questions on the complex nature of axon–glia interactions have emerged, and will shape myelin research in the future. The morphogenesis of myelin, a three-dimensional problem, is yet to be unravelled, and will require adaptation of high-resolution time-lapse imaging techniques. Myelin proteomics has revealed hundreds of new players, whose parts in myelin synthesis and maintenance remain unknown. We are only at the beginning of understanding the

dynamics of protein–protein interactions at the axon–glia interface and in mature myelin. For the CNS, the interdependence of oligodendrocytes with both axons and astrocytes needs to be explored; this may hold a key to understanding the metabolic relationship between neurons and glial cells. We have learned that oligodendrocytes and Schwann cells support survival of the axons that they ensheath. It is important to find out how they do this, as the knowledge may lead to new therapies in neuropsychiatric diseases. ■

1. Sherman, D. L. & Brophy, P. J. Mechanisms of axon ensheathment and myelin growth. *Nature Rev. Neurosci.* **6**, 683–690 (2005).
2. Jessen, K. R. & Mirsky, R. The origin and development of glial cells in peripheral nerves. *Nature Rev. Neurosci.* **6**, 671–682 (2005).
3. Simons, M. & Trotter, J. Wrapping it up: the cell biology of myelination. *Curr. Opin. Neurobiol.* **17**, 533–540 (2007).
4. Salzer, J. L., Brophy, P. J. & Peles, E. Molecular domains of myelinated axons in the peripheral nervous system. *Glia* **56**, 1532–1540 (2008).
5. Geren, B. B. & Raskind, J. Development of the fine structure of the myelin sheath in sciatic nerves of chick embryos. *Proc. Natl Acad. Sci. USA* **39**, 880–884 (1953).
   This classic work demonstrated, using electron microscopy, that myelin is a glial-cell ensheathment of axons, rather than a specialization of axons.
6. Bullock, T. H., Moore, J. K. & Fields, R. D. Evolution of myelin sheaths: both lamprey and hagfish lack myelin. *Neurosci. Lett.* **48**, 145–148 (1984).
7. Zalc, B., Goujet, D. & Colman, D. The origin of the myelination program in vertebrates. *Curr. Biol.* **18**, R511–R512 (2008).
8. Hartline, D. K. What is myelin? *Neuron Glia Biol.* **4**, 153–163 (2008).
9. Sowell, E. R. *et al.* Mapping cortical change across the human lifespan. *Nature Neurosci.* **6**, 309–315 (2003).
10. Griffin, J. W. & Thompson, W. J. Biology and pathology of nonmyelinating Schwann cells. *Glia* **56**, 1518–1531 (2008).
11. Chen, S. *et al.* Disruption of ErbB receptor signaling in adult non-myelinating Schwann cells causes progressive sensory loss. *Nature Neurosci.* **6**, 1186–1193 (2003).
    Although they were generated for a different purpose, the mutants reported in this study demonstrate that Remak cells are required for the integrity of unmyelinated C-fibre axons.
12. Nave, K. A. & Trapp, B. D. Axon–glial signaling and the glial support of axon function. *Annu. Rev. Neurosci.* **31**, 535–561 (2008).
13. Taveggia, C. *et al.* Neuregulin-1 type III determines the ensheathment fate of axons. *Neuron* **47**, 681–694 (2005).
    The first paper to show that cultured sympathetic neurons, virally transduced to overexpress NRG1, instruct associated Schwann cells to myelinate thin axons de novo.
14. Trotter, J., Karram, K. & Nishiyama, A. NG2 cells: Properties, progeny and origin. *Brain Res. Rev.* **63**, 72–82 (2010).
15. Bergles, D. E., Jabs, R. & Steinhauser, C. Neuron–glia synapses in the brain. *Brain Res. Rev.* **63**, 130–137 (2010).
16. Káradóttir, R., Hamilton, N. B., Bakiri, Y. & Attwell, D. Spiking and nonspiking classes of oligodendrocyte precursor glia in CNS white matter. *Nature Neurosci.* **11**, 450–456 (2008).
    The only report so far that some precursor cells of myelin-forming oligodendrocytes can be triggered to generate action potentials.
17. Nave, K.-A. & Salzer, J. L. Axonal regulation of myelination by neuregulin 1. *Curr. Opin. Neurobiol.* **16**, 492–500 (2006).
18. Griffiths, I. *et al.* Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science* **280**, 1610–1613 (1998).
    The first paper to show that oligodendrocytes serve the vital function of preserving the integrity and survival of axons, independent of myelin formation.
19. Edgar, J. M. *et al.* Oligodendroglial modulation of fast axonal transport in a mouse model of hereditary spastic paraplegia. *J. Cell. Biol.* **166**, 121–131 (2004).
20. Lappe-Siefke, C. *et al.* Disruption of *Cnp1* uncouples oligodendroglial functions in axonal support and myelination. *Nature Genet.* **33**, 366–374 (2003).
21. Edgar, J. M. *et al.* Early ultrastructural defects of axons and axon–glia junctions in mice lacking expression of *Cnp1. Glia* **57**, 1815–1824 (2009).
22. Yin, X. *et al.* Myelin-associated glycoprotein is a myelin signal that modulates the caliber of myelinated axons. *J. Neurosci.* **18**, 1953–1962 (1998).
23. Wang, S. S. *et al.* Functional trade-offs in white matter axonal scaling. *J. Neurosci.* **28**, 4047–4056 (2008).
24. Nave, K. A. Myelination and the trophic support of long axons. *Nature Rev. Neurosci.* **11**, 275–283 (2010).
25. de Waegh, S. M., Lee, V. M. & Brady, S. T. Local modulation of neurofilament phosphorylation, axonal caliber, and slow axonal transport by myelinating Schwann cells. *Cell* **68**, 451–463 (1992).
26. Brady, S. T. *et al.* Formation of compact myelin is required for maturation of the axonal cytoskeleton. *J. Neurosci.* **19**, 7278–7288 (1999).
27. Quarles, R. H. Myelin-associated glycoprotein (MAG): past, present and beyond. *J. Neurochem.* **100**, 1431–1448 (2007).
28. Nguyen, T. *et al.* Axonal protective effects of the myelin-associated glycoprotein. *J. Neurosci.* **29**, 630–637 (2009).

29. Simon, C. M., Jablonka, S., Ruiz, R., Tabares, L. & Sendtner, M. Ciliary neurotrophic factor-induced sprouting preserves motor function in a mouse model of mild spinal muscular atrophy. *Hum. Mol. Genet.* **19**, 973–986 (2010).

30. Keswani, S. C. *et al.* A novel endogenous erythropoietin mediated pathway prevents axonal degeneration. *Ann. Neurol.* **56**, 815–826 (2004).

31. Riethmacher, D. *et al.* Severe neuropathies in mice with targeted mutations in the ErbB3 receptor. *Nature* **389**, 725–730 (1997).
    A genetic study providing evidence that before myelination, Schwann cells are essential for the survival of neurons in dorsal-root ganglia.

32. Klugmann, M. *et al.* Assembly of CNS myelin in the absence of proteolipid protein. *Neuron* **18**, 59–70 (1997).

33. Rosenbluth, J., Nave, K. A., Mierzwa, A. & Schiff, R. Subtle myelin defects in PLP-null mice. *Glia* **54**, 172–182 (2006).

34. Werner, H. B. *et al.* Proteolipid protein is required for transport of sirtuin 2 into CNS myelin. *J. Neurosci.* **27**, 7717–7730 (2007).

35. Southwood, C. M., Peppi, M., Dryden, S., Tainsky, M. A. & Gow, A. Microtubule deacetylases, SirT2 and HDAC6, in the nervous system. *Neurochem. Res.* **32**, 187–195 (2007).

36. Li, W. *et al.* Sirtuin 2, a mammalian homolog of yeast silent information regulator-2 longevity regulator, is an oligodendroglial protein that decelerates cell differentiation through deacetylating alpha-tubulin. *J. Neurosci.* **27**, 2606–2616 (2007).

37. Zhao, S. *et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000–1004 (2010).

38. Kassmann, C. M. *et al.* Axonal loss and neuroinflammation caused by peroxisome-deficient oligodendrocytes. *Nature Genet.* **39**, 969–976 (2007).

39. Gravel, M. *et al.* 2′,3′-cyclic nucleotide 3′-phosphodiesterase: a novel RNA-binding protein that inhibits protein synthesis. *J. Neurosci. Res.* **87**, 1069–1079 (2009).

40. Lee, J., Gravel, M., Zhang, R., Thibault, P. & Braun, P. E. Process outgrowth in oligodendrocytes is mediated by CNP, a novel microtubule assembly myelin protein. *J. Cell Biol.* **170**, 661–673 (2005).

41. Readhead, C., Schneider, A., Griffiths, I. & Nave, K. A. Premature arrest of myelin formation in transgenic mice with increased proteolipid protein gene dosage. *Neuron* **12**, 583–595 (1994).

42. Kagawa, T. *et al.* Glial cell degeneration and hypomyelination caused by overexpression of myelin proteolipid protein gene. *Neuron* **13**, 427–442 (1994).

43. Simons, M. *et al.* Overexpression of the myelin proteolipid protein leads to accumulation of cholesterol and proteolipid protein in endosomes/lysosomes: implications for Pelizaeus–Merzbacher disease. *J. Cell Biol.* **157**, 327–336 (2002).

44. Anderson, T. J. *et al.* Late-onset neurodegeneration in mice with increased dosage of the proteolipid protein gene. *J. Comp. Neurol.* **394**, 506–519 (1998).

45. Ip, C. W. *et al.* Immune cells contribute to myelin degeneration and axonopathic changes in mice overexpressing proteolipid protein in oligodendrocytes. *J. Neurosci.* **26**, 8206–8216 (2006).
    The finding that a primary defect of myelinating oligodendrocytes triggers a T-cell-mediated immune response that contributes to disease severity (see also ref. 38).

46. Kroner, A., Ip, C. W., Thalhammer, J., Nave, K. A. & Martini, R. Ectopic T-cell specificity and absence of perforin and granzyme B alleviate neural damage in oligodendrocyte mutant mice. *Am. J. Pathol.* **176**, 549–555 (2010).

47. Edgar, J. M. *et al.* Demyelination and axonal preservation in a transgenic mouse model of Pelizaeus–Merzbacher disease. *EMBO Mol. Med.* **2**, 42–50 (2010).

48. Mayatepek, E., Baumann, M., Meissner, T., Hanefeld, F. & Korenke, G. C. Role of leukotrienes as indicators of the inflammatory demyelinating reaction in x-linked cerebral adrenoleukodystrophy. *J. Neurol.* **250**, 1259–1260 (2003).

49. Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502–1517 (2008).

50. Rudick, R. A. & Trapp, B. D. Gray-matter injury in multiple sclerosis. *New Engl. J. Med.* **361**, 1505–1506 (2009).

51. Trapp, B. D. & Nave, K. A. Multiple sclerosis: an immune or neurodegenerative disorder? *Annu. Rev. Neurosci.* **31**, 247–269 (2008).

52. Trapp, B. D. & Stys, P. K. Virtual hypoxia and chronic necrosis of demyelinated axons in multiple sclerosis. *Lancet Neurol.* **8**, 280–291 (2009).

53. Garbern, J. Y. Pelizaeus–Merzbacher disease: genetic and cellular pathogenesis. *Cell Mol. Life Sci.* **64**, 50–65 (2007).

54. Woodward, K. J. The molecular and cellular defects underlying Pelizaeus–Merzbacher disease. *Expert Rev. Mol. Med.* **10**, e14 (2008).

55. Uhlenberg, B. *et al.* Mutations in the gene encoding gap junction protein alpha 12 (connexin 46.6) cause Pelizaeus–Merzbacher-like disease. *Am. J. Hum. Genet.* **75**, 251–260 (2004).

56. Maglione, M. *et al.* Oligodendrocytes in mouse corpus callosum are coupled via gap junction channels formed by connexin47 and connexin32. *Glia* **58**, 1104–1117 (2010).

57. Henneke, M. *et al.* GJA12 mutations are a rare cause of Pelizaeus–Merzbacher-like disease. *Neurology* **70**, 748–754 (2008).

58. Johnson, A. B. & Brenner, M. Alexander's disease: clinical, pathologic, and genetic features. *J. Child Neurol.* **18**, 625–632 (2003).

59. van der Knaap, M. S. *et al.* Mutations in each of the five subunits of translation initiation factor eIF2B can cause leukoencephalopathy with vanishing white matter. *Ann. Neurol.* **51**, 264–270 (2002).

60. Dietrich, J. *et al.* EIF2B5 mutations compromise GFAP⁺ astrocyte generation in vanishing white matter leukodystrophy. *Nature Med.* **11**, 277–283 (2005).

61. Baes, M. & Aubourg, P. Peroxisomes, myelination, and axonal integrity in the CNS. *Neuroscientist* **15**, 367–379 (2009).

62. Nave, K. A., Sereda, M. W. & Ehrenreich, H. Mechanisms of disease: inherited demyelinating neuropathies—from basic to clinical research. *Nature Clin. Pract. Neurol.* **3**, 453–464 (2007).

63. Suter, U. & Scherer, S. S. Disease mechanisms in inherited neuropathies. *Nature Rev. Neurosci.* **4**, 714–726 (2003).

64. De Jonghe, P. *et al.* The Thr124Met mutation in the peripheral myelin protein zero (MPZ) gene is associated with a clinically distinct Charcot–Marie–Tooth phenotype. *Brain* **122**, 281–290 (1999).
    A mutation of a myelin gene is found to cause the 'axonal form' of CMT disease, which uncouples Schwann-cell functions in myelination and axonal preservation.

65. Hakak, Y. *et al.* Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc. Natl Acad. Sci. USA* **98**, 4746–4751 (2001).

66. Fields, R. D. White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* **31**, 361–370 (2008).

67. Bengtsson, S. L. *et al.* Extensive piano practicing has regionally specific effects on white matter development. *Nature Neurosci.* **8**, 1148–1150 (2005).

68. Stevens, B. & Fields, R. D. Response of Schwann cells to action potentials in development. *Science* **287**, 2267–2271 (2000).

69. Barres, B. A. & Raff, M. C. Proliferation of oligodendrocyte precursor cells depends on electrical activity in axons. *Nature* **361**, 258–260 (1993).

70. Stevens, B., Porta, S., Haak, L. L., Gallo, V. & Fields, R. D. Adenosine: a neuron–glial transmitter promoting myelination in the CNS in response to action potentials. *Neuron* **36**, 855–868 (2002).

71. Sanchez, I., Hassinger, L., Paskevich, P. A., Shine, H. D. & Nixon, R. A. Oligodendroglia regulate the regional expansion of axon caliber and local accumulation of neurofilaments during development independently of myelin formation. *J. Neurosci.* **16**, 5095–5105 (1996).

72. Sanchez, I. *et al.* Local control of neurofilament accumulation during radial growth of myelinating axons in vivo. Selective role of site-specific phosphorylation. *J. Cell Biol.* **151**, 1013–1024 (2000).

73. Bergles, D. E., Roberts, J. D., Somogyi, P. & Jahr, C. E. Glutamatergic synapses on oligodendrocyte precursor cells in the hippocampus. *Nature* **405**, 187–191 (2000).
    The discovery that unmyelinated axons make (transient) synaptic contact with the precursor cells of myelin-forming oligodendrocytes.

74. Kukley, M., Capetillo-Zarate, E. & Dietrich, D. Vesicular glutamate release from axons in white matter. *Nature Neurosci.* **10**, 311–320 (2007).

75. Káradóttir, R., Cavelier, P., Bergersen, L. H. & Attwell, D. NMDA receptors are expressed in oligodendrocytes and activated in ischaemia. *Nature* **438**, 1162–1166 (2005).

76. Micu, I. *et al.* NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* **439**, 988–992 (2006).

77. McGee, A. W., Yang, Y., Fischer, Q. S., Daw, N. W. & Strittmatter, S. M. Experience-driven plasticity of visual cortex limited by myelin and Nogo receptor. *Science* **309**, 2222–2226 (2005).

78. Budel, S. *et al.* Genetic variants of Nogo-66 receptor with possible association to schizophrenia block myelin inhibition of axon growth. *J. Neurosci.* **28**, 13161–13172 (2008).

79. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. Long-range neural coupling through synchronization with attention. *Prog. Brain Res.* **176**, 35–45 (2009).

80. Dan, Y. & Poo, M. M. Spike timing-dependent plasticity of neural circuits. *Neuron* **44**, 23–30 (2004).

81. Toonen, R. F. *et al.* Munc18-1 expression levels control synapse recovery by regulating readily releasable pool size. *Proc. Natl Acad. Sci. USA* **103**, 18332–18337 (2006).

82. Griffiths, I. *et al.* Current concepts of PLP and its role in the nervous system. *Microsc. Res. Tech.* **41**, 344–358 (1998).

83. Richardson, W. D., Kessaris, N. & Pringle, N. Oligodendrocyte wars. *Nature Rev. Neurosci.* **7**, 11–18 (2006).

84. Kukley, M., Nishiyama, A. & Dietrich, D. The fate of synaptic input to NG2 glial cells: neurons specifically downregulate transmitter release onto differentiating oligodendroglial cells. *J. Neurosci.* **30**, 8320–8331 (2010).

85. Jahn, O., Tenzer, S. & Werner, H. B. Myelin proteomics: molecular anatomy of an insulating sheath. *Mol. Neurobiol.* **40**, 55–72 (2009).

86. Saher, G. *et al.* High cholesterol level is essential for myelin membrane growth. *Nature Neurosci.* **8**, 468–475 (2005).

87. Michailov, G. V. *et al.* Axonal neuregulin-1 regulates myelin sheath thickness. *Science* **304**, 700–703 (2004).

88. Taveggia, C., Feltri, M. L. & Wrabetz, L. Signals to promote myelin formation and repair. *Nature Rev. Neurol.* **6**, 276–287 (2010).

89. Pertusa, M., Morenilla-Palao, C., Carteron, C., Viana, F. & Cabedo, H. Transcriptional control of cholesterol biosynthesis in Schwann cells by axonal neuregulin 1. *J. Biol. Chem.* **282**, 28768–28778 (2007).

90. Goebbels, S. *et al.* Elevated phosphatidylinositol 3,4,5-trisphosphate in glia triggers cell-autonomous membrane wrapping and myelination. *J. Neurosci.* **30**, 8953–8964 (2010).

91. Cotter, L. *et al.* Dlg1-PTEN interaction regulates myelin thickness to prevent damaging peripheral nerve overmyelination. *Science* **328**, 1415–1418 (2010).

92. Bremer, J. *et al.* Axonal prion protein is required for peripheral myelin maintenance. *Nature Neurosci.* **13**, 310–318 (2010).

93. Chan, J. R. *et al.* NGF controls axonal receptivity to myelination by Schwann cells or oligodendrocytes. *Neuron* **43**, 183–191 (2004).

94. Brinkmann, B. G. *et al.* Neuregulin-1/ErbB signaling serves distinct functions

in myelination of the peripheral and central nervous system. *Neuron* **59**, 581–595 (2008).

95. Carson, M. J., Behringer, R. R., Brinster, R. L. & McMorris, F. A. Insulin-like growth factor I increases brain growth and central nervous system myelination in transgenic mice. *Neuron* **10**, 729–740 (1993).

96. Ishibashi, T. *et al.* Astrocytes promote myelination in response to electrical impulses. *Neuron* **49**, 823–832 (2006).

97. Flores, A. I. *et al.* Constitutively active Akt induces enhanced myelination in the CNS. *J. Neurosci.* **28**, 7174–7183 (2008).

98. Rosenberg, S. S., Kelland, E. E., Tokar, E., De la Torre, A. R. & Chan, J. R. The geometric and spatial constraints of the microenvironment induce oligodendrocyte differentiation. *Proc. Natl Acad. Sci. USA* **105**, 14662–14667 (2008).
   This reports that cultured oligodendrocytes can 'myelinate' chemically fixed axons, which implies that only limited bidirectional signalling is required.

99. Charles, P. *et al.* Negative regulation of central nervous system myelination by polysialylated-neural cell adhesion molecule. *Proc. Natl Acad. Sci. USA* **97**, 7585–7590 (2000).

The myeloid cells of the central nervous system parenchyma

Richard M. Ransohoff[1] & Astrid E. Cardona[2]

A microglial cell is both a glial cell of the central nervous system and a mononuclear phagocyte, which belongs to the haematopoietic system and is involved in inflammatory and immune responses. As such, microglia face a challenging task. The neurons of the central nervous system cannot divide and be replenished, and therefore need to be protected against pathogens, which is a key role of the immune system, but without collateral damage. In addition, after physical injury, neural cells need restorative support, which is provided by inflammatory responses. Excessive or chronic inflammatory responses can, however, be harmful. How microglia balance these demands, and how their behaviour can be modified to ameliorate disorders of the central nervous system, is becoming clear.

The central nervous system (CNS) is now thought to consist of three main cell types: neurons, glial cells and vascular cells. In the CNS parenchyma (the tissue proper, exclusive of cerebrospinal-fluid spaces, blood vessels or meningeal coverings), microglial cells are the only type of cell that is not a neural cell or a vascular cell. Instead, microglial cells are the resident inflammatory cells of the CNS. The study of microglia provides insight into the development of the nervous system, as well as the mechanisms of neural damage and repair.

Progress has been made recently by using genetic models of development and disease, often in combination with novel imaging modalities such as two-photon microscopy. The findings have allowed researchers to begin delineating how the activation of microglia is controlled. These topics are of compelling interest to neuroscientists and other students of neurological disease, now that it is generally acknowledged that all CNS disorders are characterized by microglial-cell activation and that the progression and resolution of many diseases is contingent in part on the activity of microglia.

It is often a challenge to find strategies for modulating the underlying processes of neurodegeneration. However, numerous agents that promote or suppress aspects of inflammation or immunity are used in contemporary clinical practice as effective treatments for a range of conditions that affect virtually every organ system. Because much is known about the safety profiles and molecular mechanisms of these agents, understanding the roles of microglia in neurological diseases promises a translational reward in the near future.

Microglial-cell biology offers immunologists a lens that provides insight into neuroinflammation, neurodegeneration and neural repair. Conversely, for neurobiologists, it is essential to understand microglia in order to comprehend the inflammatory components of CNS processes. To address these topics, we review the ontogeny of microglia, how the activated state(s) of these cells is unleashed, and how some of their effector mechanisms operate in the CNS.

Microglia are the myeloid cells of the CNS parenchyma

The study of microglia is complicated by controversy and nomenclature disputes. It is reassuring to realize that this vigorous discussion about the origins and nature of microglia has been ongoing since the cells were first discovered in the late nineteenth century — and that progress has been made. At the same time, it is sobering to note that the giants who

established the techniques and concepts of neuroanatomy — including Camillo Golgi, Franz Nissl, Rudolf Virchow, Santiago Ramón y Cajal and Pío del Río-Hortega — formulated much the same questions as are being asked today.

Cells of the CNS

Early neuroanatomists endeavoured to distinguish the varied types of cell in the CNS. This was a formidable task given the dense intertwined web of cellular processes from which the cell bodies needed to be differentiated. Using Golgi's metal-impregnation techniques for staining tissue as a starting point[1,2], Cajal identified the following cell types in 1913: neurons (which he called the first element)[3]; astrocytes (the second element); and a third element (cells with small, round nuclei). Between 1919 and 1921, del Río-Hortega clarified the composition of the third element, proposing initially that it comprised two distinct populations, one of which, oligodendrocytes[4], were related more closely to astrocytes and therefore belonged to the second element. He completed his tour de force by coining the term 'microglia' for the authentic third-element cells[5], which he suggested could transform themselves into migratory phagocytic cells. These suggestions were compatible with Nissl's[6] earlier description, in 1899, of 'rod cells', which he characterized as reacting to tissue pathology and likened to the leukocytes found in other organs. In summary, microglia were accurately placed in the range of cellular elements in the CNS, as well as accorded their fundamental function as phagocytes of the nervous system, about 90 years ago.

Classification challenges

Classifying mononuclear phagocytes, the category to which microglia belong, has proved a more durable challenge than that of classifying cellular elements of the CNS. Indeed, a lively debate on this topic was published earlier this year[7]. In parallel, mononuclear phagocyte investigators and neuroscientists have been labouring to understand how microglia fit into the overall scheme of mononuclear phagocytes, both during embryonic development and in the adult CNS. This distinction between adult and embryo is important, because it helps to clarify the investigator's approach to the ontogeny, function and maintenance of microglia and their relevance to disease. Each of these processes needs to be defined at the molecular level if research scientists are to achieve

[1]Neuroinflammation Research Center (Department of Neurosciences, Lerner Research Institute) and Mellen Center for MS Treatment and Research (Neurological Institute), Cleveland Clinic, Mail Code NC30, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. [2]Department of Biology and South Texas Center for Emerging Infectious Diseases, MBT 1.216, One UTSA Circle, The University of Texas at San Antonio, San Antonio, Texas 78249, USA.
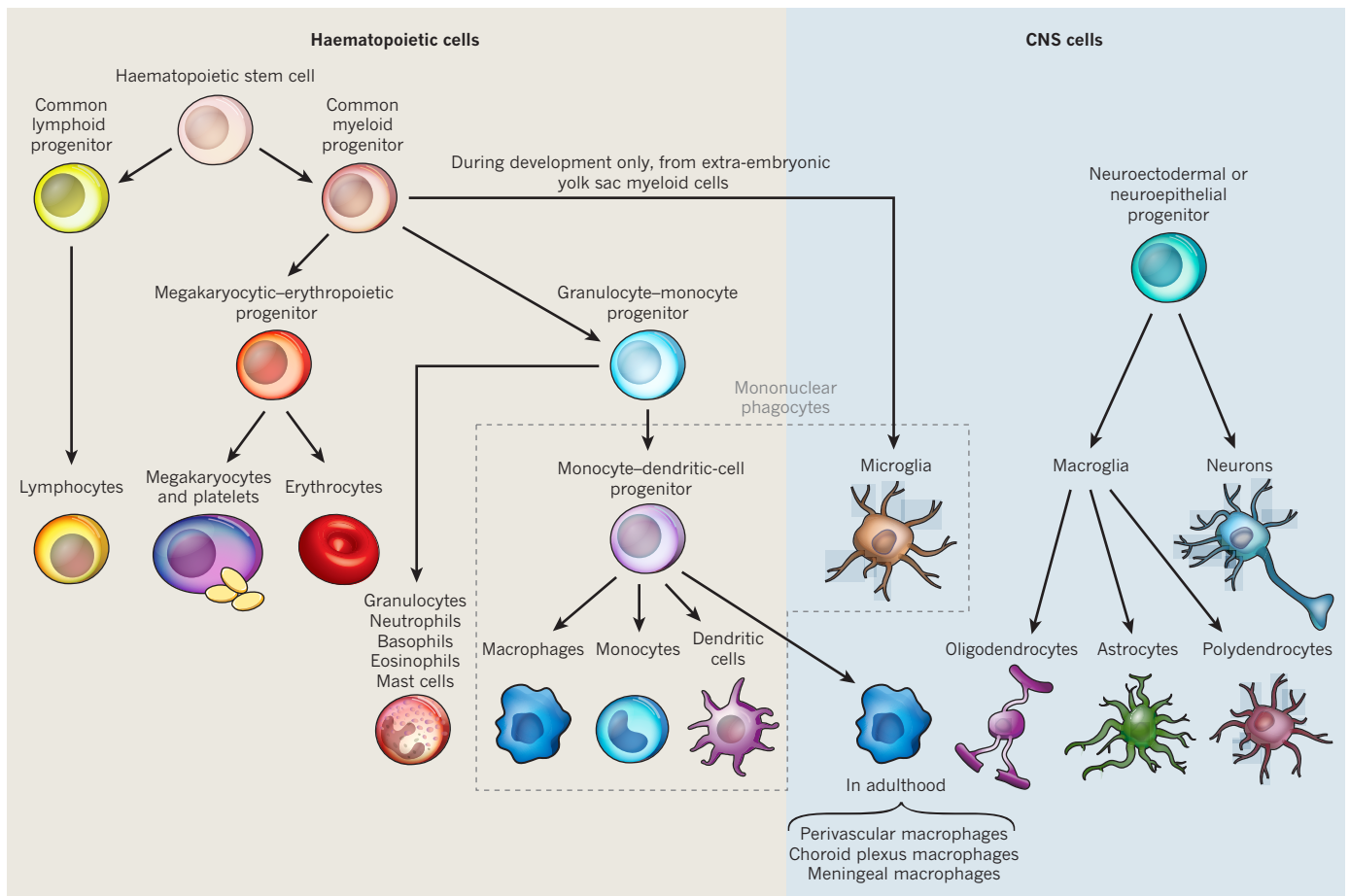
Figure 1 | How microglia are related to haematopoietic cells and CNS cells. Cells of the haematopoietic system and CNS are depicted, with arrows indicating lineage relatedness. Microglia are the only haematopoietic cells found in the parenchyma of the CNS. Note that the derivation of the perivascular, choroid plexus and meningeal macrophages from the monocyte–dendritic-cell progenitor is speculative and the blood cell from which they originate is unknown.

the goal of modulating microglial-cell properties to ameliorate CNS disorders. Furthermore, characterizing microglia and other CNS mononuclear phagocytes in this way is fundamental to understanding CNS inflammation and immunity.

Mononuclear phagocytes

Starting at the later stages of gestation, the cells of the mammalian haematopoietic system — that is, lymphocytes, erythrocytes, megakaryocytes (which produce platelets), granulocytes and mononuclear phagocytes — are renewed continually from bone-marrow haematopoietic stem cells throughout life (Fig. 1). Cells of the mononuclear phagocyte lineage are distributed across the bone marrow, blood and tissues, and include bone-marrow mononuclear phagocyte progenitors, blood monocytes, tissue macrophages and dendritic cells. These cells cannot be distinguished in a straightforward manner by immunohistochemical staining of tissues or immunofluorescence flow cytometry using cells from blood or tissue lysates. Instead, mononuclear phagocytes are classified by a combination of phenotype (cell-surface or enzymatic markers), location (bone marrow, blood or tissue), function and derivation (cell of origin).

Dendritic cells (DCs) are noteworthy because they are uniquely required for initiating adaptive immune responses, by taking up antigenic material in the tissue, gaining access to the local lymph nodes, and presenting antigen efficiently to naive T cells there[8]. The current consensus is that DCs can be found in the spleen, lymph nodes and skin, and at mucosal surfaces, but are unlikely to be present in healthy CNS tissues. Indeed, the cellular basis of CNS immune privilege rests mainly on the absence of dendritic cells in the healthy CNS[9].

During postnatal life, myeloid progenitor cells in the bone marrow

give rise to common monocyte–dendritic-cell progenitors, which in turn yield blood monocytes and dendritic-cell progenitors. Dendritic cells arise from this committed progenitor. It remains uncertain whether monocytes differentiate into dendritic cells *in vivo,* although dendritic cells are readily obtained *in vitro* from monocytes by using well-established protocols. Tissue macrophages can be derived from blood monocytes or from circulating mononuclear phagocyte progenitors that are incompletely defined; they are maintained either by local self-renewal or by influx of cells from the circulation.

Microglia are derived from myeloid-lineage progenitors

Controversy about the possible origins of microglia from neuroectodermal cells, instead of myeloid progenitors, persisted well into the 1990s[10,11]. This debate occurred despite the fact that, by 1921, del Río-Hortega had clarified that microglia constituted Cajal's 'third element' of CNS cells. Furthermore, he had proposed that microglia were mesodermal in origin, not ectodermal, a concept that is regarded as correct at present. Perhaps the most convincing data favouring the myeloid origin of microglia came with the demonstration that differentiation of cells in the myeloid lineage, including into macrophages and neutrophils, failed in mice that lacked the transcription factor PU.1 (*Pu.1*$^{-/-}$ mice) and that no microglia were detected in the CNS of these mice[12]. It is worth noting that yolk-sac embryonic stem cells from *Pu.1*$^{-/-}$ mice can yield macrophage-like cells *in vitro* and that at embryonic day (E) 11.5 *Pu.1*$^{-/-}$ mice had typical embryonic phagocytes. These findings suggest that myeloid progenitors from the yolk sac may be a different population to those arising from definitive haematopoiesis in the embryo[13]. Interestingly, embryonic phagocytes appeared in *Pu.1*$^{-/-}$ mice several days after microglial progenitors are found in

wild-type animals, underlining the restricted time frame within which microglial progenitors seed the CNS[14].

Earlier well-designed studies assigned some parenchymal microglial cells to the neuroectodermal lineage for reasons that were logically derived from careful observations, although this interpretation is now regarded as incorrect. In one study[15], myeloid-marker-bearing microglia were meticulously tracked in the CNS anlage of rodents during early embryonic life, but myeloid-marker-bearing cells were not detected later in gestation, leading to the plausible interpretation that this wave of early infiltrating cells underwent programmed cell death. In addition, the researchers transferred bone-marrow cells that had been labelled with a genetic marker to irradiated rodents. Subsequent assessment of the CNS tissues showed that ~10% of parenchymal microglia bore the bone-marrow-derived genetic marker. Taken together, these results were reasonably construed to mean that after the widespread death of early-embryonic microglia, their replacements came from a non-bone-marrow (neuroectodermal) source. With the advantage of hindsight, it seems more probable that the preliminary wave of microglial-cell colonizers downregulated their myeloid markers during maturation in the profoundly immune-inhibitory CNS environment, and then weakly expressed selected indicators of their myeloid provenance only later[16]. A recent report used lineage tracing to show conclusively that adult mouse microglia are derived from primitive myeloid precursors that arise in the extra-embryonic yolk sac in a remarkably restricted time frame, just before E8.0, and invade the embryo after blood vessels form, at E9.0. There seem to be no contributions to microglia from haematopoiesis in the embryo or the adult. Microglia are accordingly distinguished from other tissue macrophages by their properties of local self-renewal and resistance to ionizing radiation[14]. Given these results, it was not surprising that other hypotheses, such as that microglia arose from non-endothelial vessel-associated mesodermal cells, have failed to gain experimental support[8].

Microglia as mononuclear phagocytes

Taking into account the findings described above, in this Review we use the following definitions. Monocytes are mononuclear phagocytes circulating in the blood. Macrophages are phagocytic cells that reside in the tissue. Dendritic cells are mononuclear phagocytes that initiate immune responses by stimulating T cells[17], and are present in the spleen, lymph nodes and skin and at the mucosal surfaces. Microglia are myeloid cells of the CNS parenchyma.

Microglia have the special distinction of being both glia of the CNS and a unique type of mononuclear phagocyte (Fig. 1). In addition to the microglia, which are present in the CNS parenchyma, the CNS (including the brain and spinal cord, as well as their meningeal coverings, the subarachnoid and perivascular spaces and the choroid plexus) contains other mononuclear phagocytes: meningeal macrophages, choroid plexus macrophages, epiplexus cells and perivascular macrophages, each with a distinct location, marker profile and presumed function in host defence (Table 1).

These limited definitions of the mononuclear phagocyte subgroups are offered in the hope that they will be useful until definitive molecular markers for these populations emerge. In this regard, microglia are recognized at present as a subset of tissue macrophages that do not have dendritic-cell characteristics *in vivo*.

Some provisional conclusions can be drawn from this definition of microglia. First, if microglia do not initiate immune responses by activating naive T cells through antigen presentation (as dendritic cells do), then the lack of dendritic cells residing in the CNS parenchyma is a satisfactory cellular basis for the immune privilege of the CNS[9]. Second, it remains important to determine the capacity of microglia to restimulate tissue-invading memory T cells during host defence or immunopathological responses such as those that give rise to multiple sclerosis. Finally, it is incumbent on researchers to apply their understanding of mononuclear phagocyte biology to clarifying the consequences of the unique origins and maintenance of microglia throughout embryonic and adult life. We propose that the capacities of microglia are defined in part by their ontogeny and that monocyte-derived cells that enter the CNS during disease have properties that are distinct from those of microglia, as suggested by studies in animal models[18–20]. If this hypothesis is correct, then exploiting our new understanding of microglial-cell ontogeny[14] will help researchers to decide when to modulate microglial-cell activation, when to block infiltrating monocytes or their progenitors, and what, if any, potential exists for delivering genes or gene products to the CNS, in the context of neurological disease, by loading them onto potential microglial-cell progenitors[21].

Ontogeny of microglia

In summary, microglia are mononuclear phagocytes, and we are now in a position to consider when their progenitors enter the CNS and which route of entry they use[14].

Table 1 | Location and immune function of mononuclear phagocytes in the CNS

Type	Location	Phenotype	Immune function	Outcome of interaction with T cells
Choroid plexus macrophages	Choroid plexus stroma	Dendritic-cell-like: armed with both MHC class II molecules and co-stimulatory molecules, allowing antigen presentation	Present antigen derived from the peripheral circulation to T cells, thereby restimulating memory CD4[+] T cells as they traffic into the cerebrospinal fluid	Interact with T cells that are crossing from blood to CSF Host defence against extracellular bacteria Possibly autoimmune responses Facilitation of activation of parenchymal vasculature of the blood–brain barrier
Meningeal macrophages	Associated with the pial vasculature and the stroma of the subarachnoid space	Armed mainly with MHC class II molecules	Present peptide antigens derived from breakdown of CNS cellular components (cellular or pathogen derived) to T cells, thereby restimulating memory CD4[+] T cells as they traffic through the cerebrospinal fluid	Interact with CSF T cells in subarachnoid space Host defence against viruses and tumours Possibly autoimmune responses Facilitation of activation of parenchymal vasculature of the blood–brain barrier
Perivascular macrophages	Within the perivascular spaces of the parenchymal and subpial vessels on either the 'vascular' side or the 'parenchymal' side of the glial limitans basement membrane	Armed with both MHC class II molecules and co-stimulatory molecules, allowing antigen presentation	Restimulate effector T cells that have extravasated across activated subpial vessels (which have interendothelial tight junctions) or the parenchymal blood–brain barrier	Interact with T cells in parenchymal perivascular space and facilitate invasion of the parenchyma proper during the effector phases of host defence and autoimmunity
Parenchymal microglia	Within the parenchyma near the neurovascular unit but not directly associated with vessels May form part of the perivascular glial limitans	Express cytosolic more than membrane MHC class II	Local source of potential phagocytes Produce antibacterial effectors through an oxidative burst Produce cytokines	Interact with T cells in the parenchyma proper Much interaction occurs at a distance, through response to T-cell-secreted cytokines such as IFN-γ, and is not antigen specific
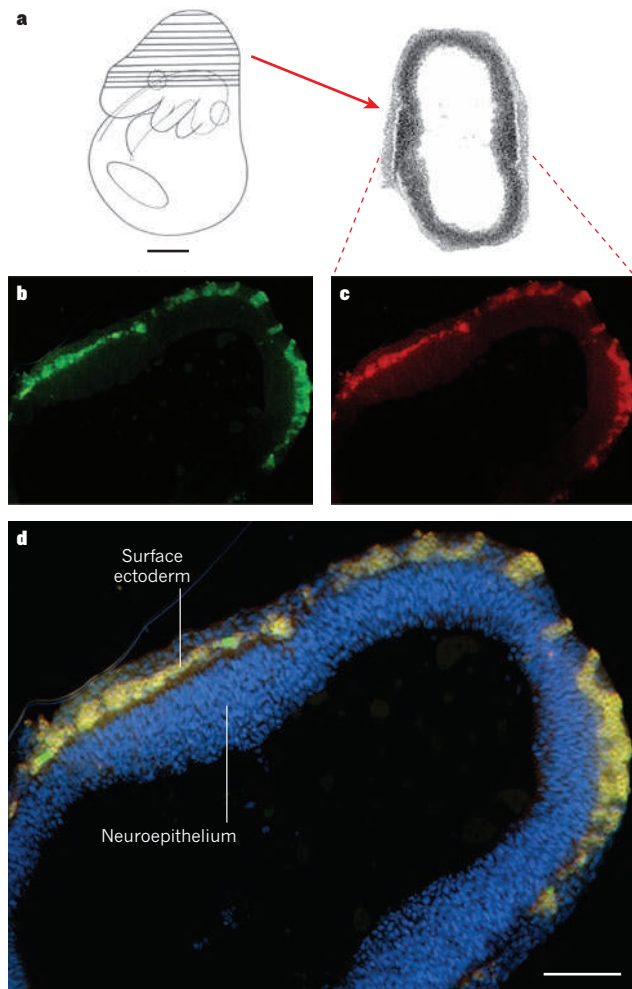
Figure 2 | Microglial precursor cells approach the CNS during embryonic development. To study the origin of microglia in the CNS, we assessed developing CNS tissues for the expression of chemokine receptors that are present at the surface of mononuclear phagocytes. Mice expressing green fluorescent protein (GFP) under control of the *Cx3cr1* promoter (denoted CX₃CR1-GFP⁺) and red fluorescent protein (RFP) under control of the *Ccr2* promoter (CCR2-RFP⁺) were mated. All mononuclear phagocytes express one or both of these receptors. Tissues from the resultant offspring were analysed *in vitro* at various time points during gestation. Sections were stained with DAPI (4′,6-diamidino-2-phenylindole) to visualize cell nuclei (blue). In addition, confocal fluorescence microscopy images were acquired, and 20-μm *z*-stacks show the developing neuroepithelium. **a**, The embryo is shown in schematic form on the left, with solid lines indicating the plane of the section studied. (Schematic and bright-field micrograph (right) reproduced, with permission, from ref. 98.) **b–d**, Confocal image of a representative section showing the distribution of CX₃CR1-GFP⁺ cells (**b**), CCR2-RFP⁺ cells (**c**) and a merged image of **b** and **c** (**d**) at E9.5. In the merged image, a dense band of mononuclear phagocytes (yellow) is observed between the neuroepithelium and the surface ectoderm. Distribution of GFP⁺ and RFP⁺ cells shows that both receptors are expressed by mononuclear phagocytes localized between these tissue layers. Moreover, from these data, it is clear that microglial precursor cells approach the neuroepithelium at a time before the onset of fetal haematopoiesis or vascularization of the developing CNS. Scale bars, 500 μm (**a**) and 100 μm (**d**).

Microglia colonize the embryonic CNS

As noted above, cells bearing myeloid markers are found immediately adjacent to the developing CNS of rodents as early as ~E8.5 (Fig. 2 shows E9.5), and they appear within the parenchymal anlage as early as E10.5 (Fig. 3). The myeloid progenitors of the yolk sac are the source for the embryonic wave of myeloid cells that colonize the developing telencephalon and brain stem, and presumably the spinal cord as well.

It seems most plausible at present that all cells destined to give rise to parenchymal microglia share origins and developmental pathways[22]. Because macrophage-like cells from the yolk sac are functionally distinct from other mononuclear phagocytes, the properties of microglia will probably need to be investigated by focusing specifically on microglia themselves[13]. In the first few days after birth, a marked increase in the number of microglia is observed within rodent CNS tissues[23]. This remarkable expansion of the microglial population arises from *in situ* proliferation. The migration cues for microglial-progenitor-cell transit across vascular beds remain obscure, because microglia are found in the appropriate numbers in the CNS of mice lacking crucial regulators of leukocyte migration[24]: molecules that aid leukocyte interaction with the endothelium such as intercellular adhesion molecule 1 (ICAM1), chemoattractant cytokines (chemokines) such as CCL2, CCL5 and CX₃CL1, and chemokine receptors such as CCR1, CCR2, CCR5 and CX₃CR1.

Questions raised

One problematic issue is immediately apparent from this discussion: are there microglial subpopulations whose attributes are determined in part by their localization within the CNS? At present, the only conclusion that can be made with confidence is that researchers do not yet have the appropriate techniques to uncover the detailed relationships between the haematopoietic cells that enter the various regions of the CNS during development and the microglia found there in adult life.

The definitions of stages along the branched, interrelated myeloid differentiation pathways[7] rely mainly on surface and enzymatic markers. About 50 surface markers characteristic of the mononuclear phagocyte lineage can be readily listed, of which perhaps 35 are thought to be useful for discriminating subsets of macrophages and dendritic cells[8]. In medicine and science, philosophical chasms divide 'lumpers' from 'splitters'[25]. A member of the former camp has noted that the "number of subpopulations of mononuclear phagocytes one can define is an exponential function of the number of markers examined"[7]. This complexity synergizes with differences in temperament, scientific approach and research priority among investigators, limiting progress towards achieving a consensus about subsets of microglia and the possibility that there may be distinct varieties of activated microglia.

Do embryonic microglia carry out developmental functions?

Many other questions remain open. One of these is, what are the key functions of microglia that drive their infiltration of the CNS during development? Numerous important activities have been proposed: directing the invading vasculature[26], removing apoptotic cells (which the developing CNS generates in profusion)[27,28] and synaptic pruning[29]. Elegant histological studies have shown that microglia ingest whole cells (not just cellular processes) and are often, but not always, spatiotemporally concentrated with high numbers of apoptotic cells[30–32]. To interpret these data, it is important to consider that cells other than mononuclear phagocytes participate in the removal of apoptotic cells from the developing CNS[33]. Furthermore, the molecules involved in the recognition and clearance of apoptotic cells are ancient and well conserved throughout evolution, so the expression of these by non-microglial cells is unsurprising[34]. Microglia may help to initiate apoptosis of excess cells generated during development[35,36], although this concept has not been universally accepted[28]. These processes have been difficult to investigate because it is challenging to remove or disable microglia experimentally during development. Recent results showing that colony stimulating factor 1 receptor and, potentially, interleukin 34, are required for the appearance of microglia in the CNS[14] may provide tools for addressing some of these issues.

If it were confirmed that microglia are involved in eliminating excess cells (both by stimulating programmed cell death and removing debris) and trimming away unneeded synaptic connections, this could implicate embryonic microglia in establishing the intricate circuitry of the developing CNS. However, *Pu.1⁻/⁻* mice, which are deficient

in microglia, seem to behave normally until their death from infection at about day 17 after birth despite antibiotic treatment, suggesting that microglial-cell contributions to prenatal neural development are subtle.

Abnormal microglia alter CNS physiology

Nevertheless, minor abnormalities in the microglial-cell input to neural function can combine to have large effects. One example from medicine is Nasu–Hakola disease, a rare autosomal recessive genetic disorder[37,38]. Individuals with this condition have bone cysts and dementia in early adulthood. Its genetic basis lies in homozygous null mutations of the gene encoding TREM2 (triggering receptor expressed on myeloid cells 2). Absence of this cell-surface protein (which is present only on microglia among the cells of the CNS) causes selective functional abnormalities of microglia and of osteoclasts (cells that break down bone and are also derived from the mononuclear phagocyte lineage). Remarkably, a disease phenocopy is observed in individuals lacking the transmembrane adaptor signalling protein DAP12 (refs 38, 39), which transduces signals from TREM2, as well as from other cell-surface components of haematopoietic cells[37]. Mice lacking TREM2 as a result of a targeted gene deletion show some features of the human disorder[40]. Given that a novel loss-of-function mutation in *TREM2* caused early-onset dementia but not bone cysts in one human family[41], disruption of TREM2 is the most persuasive argument that abnormal microglial-cell physiology can disturb homeostasis in the CNS. On the basis of *in vitro* studies using neural cell lines, cell-surface heat-shock protein 60 (HSP60) has been proposed to be a ligand for TREM2 (ref. 42).

Magnetic resonance imaging shows brain abnormalities in individuals with Nasu–Hakola syndrome[39]. A genetic abnormality specific to microglia can also severely perturb behaviour without causing obvious neuroanatomical defects. In particular, mice lacking the homeobox gene *Hoxb8* do not have neuroanatomical defects but have pathological grooming behaviour, in a pattern with similarities to obsessive–compulsive disorder in humans[43]. Because the Hox genes are widely expressed and involved in organ patterning, it was unsurprising to find Hoxb8 expression both in spinal-cord neurons and in brain microglia. The selective deletion of Hoxb8 from spinal-cord neurons by conditional gene targeting caused defective sensory signalling but unexpectedly did not correlate with the obsessive grooming behaviour[44]. Surprisingly, replacement of the haematopoietic cells of these mice with wild-type haematopoietic cells rescued the defect in grooming without altering the sensory abnormalities, thereby dissociating the two phenotypes of sensory impairment and obsessive grooming[44]. Furthermore, selective deletion of *Hoxb8* from haematopoietic cells during development recapitulated the aberrant grooming behaviour in adult mice. Neither T cells nor B cells were required for obsessive grooming in mice lacking haematopoietic-cell *Hoxb8* (ref. 44). The researchers concluded that *Hoxb8* expression in microglia underpins the grooming defect[44]. Interestingly, *Hoxb8* mRNA expression (tracked in otherwise-wild-type mice, by placing a genetic reporter under the control of the *Hoxb8* promoter) was present in about 40% of microglia, and reached that level during the first two weeks of postnatal life.

Although it is risky to make conclusions on the basis of one marker (perhaps *Hoxb8* was simply upregulated during this postnatal period), it is possible that an identifiable subpopulation of microglia regulate the neural substrate in terms of grooming behaviour. The obsessive grooming seen in mice lacking the *Hoxb8* gene in microglia was not associated with marked abnormalities in neural cell populations. Instead, aberrant interactions of microglia with neurons, either at a distance through secreted factors[45] or by direct contact with synapses[46], might be responsible for the behavioural phenotype. Finally, it should also be noted that clinical non-CNS haematopoietic disorders such as lymphoma are associated with severe itching, which could mimic obsessive grooming. Because the bone-marrow chimaerism replaced the entire blood system with *Hoxb8*-sufficient cells, microglia may not provide the full explanation for the phenotype.

Results of microglial–cell responses can be difficult to specify

Microglia show morphological indications of activation in the brains of patients dying as a result of neurodegenerative diseases such as Alzheimer's disease. *In vitro* studies clearly show that cultured microglia can be activated by released amyloidogenic peptides (termed amyloid-β peptides) derived from β-amyloid precursor protein. Proteolytic generation of these peptides is thought to be an early event in the pathogenesis of Alzheimer's disease. The peptides can form toxic soluble oligomers, and they also constitute the fibrils that form amyloid plaques, a major histological hallmark of Alzheimer's disease. The response of microglia to soluble and fibrillar amyloid-β peptides has been studied intensively since it was proposed that phagocytosis of these peptides by microglia could ameliorate the pathogenic cascade that occurs in the brains of individuals with Alzheimer's disease. Conversely, the release of pro-inflammatory mediators by activated microglia could promote Alzheimer's disease pathology. The current understanding of receptor-mediated recognition of aggregated (fibrillar) amyloid-β by microglia[47] is that the signalling leading to phagocytosis and the signalling resulting in production of reactive oxygen and/or nitrogen species might be indistinguishable. Therefore, the acute response of microglia to fibrillar amyloid-β peptides, an endogenous indicator of CNS tissue disturbance, might carry inseparable beneficial and harmful consequences. Investigators have made heroic efforts to deconvolute how microglia respond to fibrillar and soluble amyloid-β peptides, but the results have been challenging to interpret, given their apparent inconsistencies. Complexity is
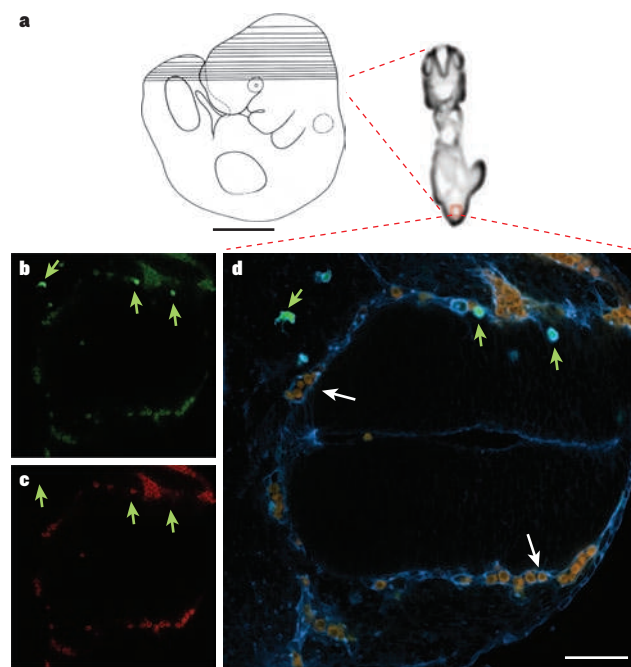
Figure 3 | Microglia colonize the CNS during embryonic development. The study was carried out as described in Fig. 2, except that axial sections of embryos were stained with isolectin IB4 (blue) to visualize the vasculature and microglia. **a**, The embryo is shown in schematic form with solid lines indicating the planes of the sections studied (left) and an image of the plane depicted in **b**–**d** (right) with a square indicating the selected section (the developing spinal cord). (Reproduced, with permission, from ref. 98.) **b**–**d**, Distribution of CX₃CR1-GFP⁺ cells (**b**), CCR2-RFP⁺ cells (**c**) and a merged image of **b** and **c** (**d**) at E10.5. CX₃CR1-GFP (**b**) but not CCR2-RFP (**c**) is expressed by developing microglia. In the merged image (**d**), newly migrated amoeboid microglia (green arrows) appear CX₃CR1-GFPʰⁱ, IB4⁺ and CCR2-RFP⁻. Mononuclear phagocytes remaining in the vasculature (white arrows) have a CX₃CR1-GFPˡᵒʷCCR2-RFP⁺ phenotype. From these data, it is clear that amoeboid microglia enter the developing CNS (parenchymal anlage) around E10.5, coincident with vascularization. Scale bars, 1 mm (**a**) and 100 μm (**d**).

inherent in the biology: signalling downstream of the recognition of fibrillar amyloid-β peptides by microglia involves an aggregate comprising various combinations of nine transmembrane and endosomal proteins, including class A and B scavenger receptors, $\alpha_6\beta_1$-integrin, CD14, CD47, Toll-like receptor 2 (TLR2), TLR4, TLR6 and TLR9 (ref. 48). In addition, to some extent, inconsistent data arise from using a variety of extensively manipulated *in vivo* genetic models. The issues have been incisively discussed recently[48] in a review that also underlined the importance of understanding whether microglial-cell activation is essential for the success of passive and/or active immunotherapy for individuals with Alzheimer's disease. Finally, data from these varied experiments highlight the different capacities of resident microglia and infiltrating blood-derived macrophages for clearing amyloid plaques[18,19].

Microglial–cell homeostasis in the adult CNS

In the adult CNS, more than 10% of all cells are microglia, varying from about 5% to nearly 20% depending on the location in the neuraxis. In the healthy CNS, microglia have a distinctive distribution, with radial, non-overlapping processes, a small cell soma and each cell appearing to occupy its own domain.

Surveillant not 'resting' microglia

Microglia of the healthy CNS have, for decades, been termed 'resting' to discriminate their morphology from the microglia of the injured or diseased CNS. Two-photon imaging of the cerebral cortex of live mice[49] demonstrated vividly that microglia continually palpate the environment with their processes, and it has been estimated[50] that the entire volume of the brain is examined by these endlessly busy cells approximately every four to five hours. The term 'surveillant' has been advanced to describe how microglia continuously monitor the healthy neural tissue. Although we salute the elegance and beauty of these landmark imaging studies, it must be recognized that the information comes exclusively from the outer layers of the cerebral cortex, that the densities and microenvironments of microglia differ throughout the neuraxis, and that the surveillant behaviour of microglia might also vary depending on their location within the CNS. In this regard, recent reports of the feat of sophisticated two-photon imaging of mouse spinal-cord microglia are most welcome[51,52].

Microgliosis indicates increased microglia numbers

In CNS pathology, an increase in the number of microglia is almost always observed. Given that many clinical CNS disorders are associated with disrupted function of the blood–brain barrier, allowing the entry of cells from the bloodstream, it has been unclear to what extent microglia proliferate *in situ* after injury and acquire the appearance of macrophages, and conversely, how many monocytes enter the CNS from the bloodstream and subsequently adopt microglial-cell morphology[53]. This question has proven surprisingly contentious[16,54] and is particularly difficult to settle definitively because no known cell-surface or enzymatic markers differentiate the activated macrophage-like progeny of microglia from those of blood monocytes. Furthermore, the technique that is most widely used for these studies is radiation bone-marrow chimaerism, in which animals are sublethally irradiated and then their bone marrow is replaced with genetically labelled haematopoietic stem cells; microglia remain because they are resistant to radiation. This method is fraught with unavoidable confounding factors[54–56]. At present, it seems most plausible that the number of microglia increases through proliferation *in situ* in pathological states[55] in which the blood–brain barrier has not been grossly disrupted, in such a way as to allow entry of circulating cells to the CNS. One well-studied example comes from transection of the facial nerve in the cheek, resulting in brisk accumulation of microglia around the axotomized ipsilateral brainstem nucleus of the facial nerve[57]. In previous experiments, results from using bone-marrow chimaerism suggested that microgliosis (accumulation of an abundance of microglia in a focal site of injury) stemmed from large-scale infiltration

of blood monocytes that had differentiated into microglia. However, it has recently become clear that several artefacts complicate the interpretation of these studies. In particular, a course of cranial irradiation permanently alters the cerebral vasculature; the transfer of bone-marrow cells (taken by femoral flushing) introduces non-physiological levels of haematopoietic stem cells into the circulation; and damage to the CNS is also required to facilitate CNS invasion by circulating monocytes[54–56]. Given these findings, it seems likely that the complement of microglia in the CNS parenchyma is established during the early-embryonic period and is self-renewing thereafter[54,55].

Removal of inhibition as the first step in microglial–cell activation

Microglia in the diseased or damaged CNS are termed 'activated' and are easily distinguished by conventional tissue histology because they bear little morphological resemblance to those found in the healthy brain.

From morphology to molecules

On the basis of their extensively branched processes, surveillant microglia in CNS tissue sections from healthy animals are designated 'ramified'. The morphology of 'activated' microglia includes enlargement of the soma, retraction and shortening of processes and increased expression of myeloid cell markers. Mechanistic information about this morphological transformation by microglia has been obtained: using two-photon imaging as an analytical tool, it was shown[49] that microglia initially extend processes in response to a focal laser lesion and that ATP (either released by damaged cells or in a regulated manner by astrocytes) governs extension of the processes, which rapidly wall off small destructive lesions. The purinoceptor that underlies this response is $P2Y_{12}$ (ref. 58). Subsequent morphological changes were found to be associated with downregulation of expression of $P2Y_{12}$, conversion of ATP to adenosine by the microglial-cell ectoenzymes CD73 and CD39, and increased expression of adenosine A2A receptor[59], which mediated process retraction.

At the far end of the range of morphological activation, microglia are termed 'amoeboid', indicating rounded cells with sparse processes and capable of phagocytosis. Interestingly, this morphological transformation recapitulates in reverse that which transpires during development, with amoeboid microglial progenitor cells entering the CNS and slowly adopting the ramified morphology of mature microglia.

The concept that activation of microglia occurs along a linear range, as suggested by the descriptive morphology of the cells in a variety of diseases and models, has been discarded[60]. Current concepts of activation arise, in part, from research into monocyte and macrophage biology, for which it has been elegantly and forcefully demonstrated that differing stimuli cause differentiation along distinct pathways of gene expression. Macrophages that have been exposed to different stimuli are designated M1, M2a, M2b and M2c macrophages, depending on the stimulus and the context. Functional significance is attributed to these varied states of activation, with, for example, M1 macrophages thought to specialize in pathogen elimination and the various categories of M2 macrophage possibly involved in tissue remodelling and repair[7,61]. M1 macrophages and the varied M2 subtypes can frequently appear similar but are distinguished from each other by defined profiles of stimulus-induced gene expression and effector function, lending credence to the categorization. Consistent with this concept, morphologically identical microglia in varied models of CNS disease show widely divergent patterns of gene expression.

It has not, thus far, been possible to develop a coherent account of the different activation profiles of microglia in a way that points researchers towards a usable terminology to differentiate varied microglial activation programs (which might be termed 'Mi1', 'Mi2' and so on). Nevertheless, it is considered highly likely that microglia behave similarly to macrophages, by polarizing their activation states to achieve effector responses that are appropriate for varying challenges to the CNS[16]. One striking example is termed 'stress conditioning': in this

process, a precisely timed and calibrated peripheral challenge to the organism induces a response from microglia that lessens the impact of a subsequent CNS insult. Examples of such a peripheral challenge are endotoxin injection or altered body temperature, whereas the most-studied CNS insults include seizure and stroke[62]. Microglia have been proposed as passive targets of neuroprotective preconditioning, with many experiments showing dampened inflammatory responses. However, genetic ablation studies have shown that endotoxin-induced protection against seizures is eliminated in the absence of microglia, suggesting that microglia have an active role in this form of 'beneficial' inflammatory response[63].

The varied products and activities of activated microglia have been catalogued and discussed over the past few decades[16,64–67]. Much of this work was carried out using *in vitro* cultures, which have disadvantages, including that serum (derived from plasma to which microglia in the healthy brain are never exposed) is commonly used[16]. Detailed gene-expression analysis[68] confirmed a long-standing wariness about extrapolating from such cultures to the *in vivo* situation. The investigators showed, fascinatingly, that microglial-cell cultures that had been stimulated with cytokines and bacterial cell-wall components expressed genes in a pattern that was more similar to that of macrophages from the abdominal cavity (challenged with the same stimuli) than to those of brain microglia (isolated from animals into which the same cocktail of cytokines and bacterial components had been injected into the CNS). The *coup de grâce* came from the demonstration that macrophages that infiltrated the brain after these injections expressed genes more akin to those expressed by brain microglia than those induced in cultured macrophages. The salient conclusion from the study[68] was that the CNS environment contributed more to gene expression by mononuclear phagocytes than did the stage of lineage differentiation. As a generalization, the study of microglial-cell cultures indicates what these cells can do but might not accurately reflect their activities *in vivo*.

On activation, microglia can produce numerous protein mediators, including those categorized as cytokines (both pro-inflammatory and anti-inflammatory cytokines), growth factors, chemokines and neurotrophins. As mononuclear phagocytes, they can also generate reactive oxygen and nitrogen species through enzymes involved in the oxidative burst mechanism and inducible nitric oxide synthase. Microglia can be phagocytic. In addition, they can express molecules that are associated with an ability to stimulate T cells with antigen (a feature that is typical of antigen-presenting cells).

Repressed microglia of the healthy CNS

Researchers spent decades before conclusively defining microglia as myeloid cells from the haematopoietic lineage. One reason for the prolonged controversy over this topic was that microglia do not express many of the markers of mature myeloid cells in other organs. It has been suggested that microglia are restrained by numerous microenvironmental inhibitory influences , many of which are produced by neurons[16,69] (Fig. 4). Neuron–microglial-cell inhibitory signalling is mediated by the following interactions: CD200–CD200 receptor (CD200R)[70], CD22–CD45 (also known as PTPRC)[71], CD172A (also known as SIRP-α)–CD47 (refs 71, 72) and CX$_3$CL1–CX$_3$CR1 (refs 73, 74). Some of these signals (CD22–CD45 and CD200–CD200R) require cell–cell contact. Signalling from neuronal CX$_3$CL1 to microglial-cell CX$_3$CR1, and from neuronal ICAM5 to microglial-cell LFA1 (lymphocyte function associated antigen 1; also known as CD11a–CD18 and α$_L$β$_2$-integrin), can occur through contact between cell-surface receptors and cell-surface-bound ligands, or at a distance after the generation of the soluble ligands CX$_3$CL1 or ICAM5, respectively[75–77]. In the healthy CNS, two-photon imaging shows that microglia 'inspect' neuronal synapses *in vivo* in the cerebral cortex of adult rodents[46]. It is not clear whether the subcellular distribution of microglial-cell CD200R and neuronal CD200 or other contact-dependent components would accommodate contacts in the manner demonstrated in this study.
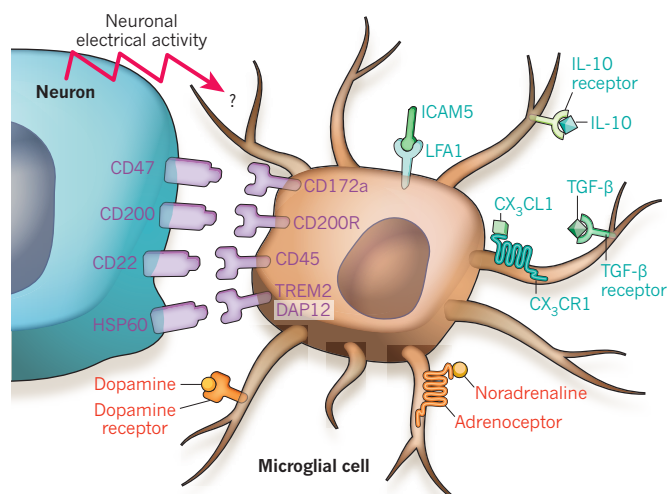
Figure 4 | Neuronal inhibitory influences on parenchymal microglia. Neurons inhibit microglia in several ways. Contact-dependent inhibitory influences are shown in purple. Soluble inhibitory cytokines or adhesion molecules and their receptors are shown in blue, and soluble neurotransmission-associated inhibitors and their receptors are shown in red. For further information, see the following references: contact-dependent inhibition, through CD200–CD200 receptor (CD200R)[70,99], HSP60–TREM2–DAP12 (refs 39–42, 100, 101), CD22–CD45 (ref. 102) and CD47–CD172A[71,72]; cytokine- or soluble adhesion-molecule-mediated inhibition, through soluble ICAM5 (intercellular adhesion molecule 5)–LFA1 (lymphocyte function associated antigen 1)[77], CX$_3$CL1–CX$_3$CR1 (refs 80, 103), interleukin 10 (IL-10)–IL-10 receptor[104] and transforming growth factor-β (TGF-β)–TGF-β receptor[104,105]; neurotransmission-associated inhibition through dopamine–dopamine receptor (D1, D2, D3 and D4 subtypes)[106], and noradrenaline–adrenoceptor[106]; and neuronal electrical-activity-mediated inhibition[107].

Unleashing microglia by removing inhibition

An interesting corollary of this precept that neuronal inputs restrain microglia is that microglial-cell activation may be equivalent, to some extent, to removing inhibition. This hypothesis has the attractive feature of introducing a teleological explanation for the abrupt response of microglia to neuronal injury: specifically, that microglia sense damage when the neuronal expression of inhibitors is silenced. Consistent with this idea, microglia are unusual among all mononuclear phagocytes in that they express ion channels, neurotransmitter receptors and an atypically rich range of purinoceptors[69,78,79].

The relevance of CX$_3$CR1 regulation of microglial-cell activation *in vivo* was demonstrated by studying toxin-induced neuron injury and genetic models of neurodegeneration[80]. In these models, *Cx3cr1*$^{-/-}$ mice showed exacerbated neuron loss compared with *Cx3cr1*$^{+/+}$ mice. This result could be attributed to altered microglial-cell activation, because, among CNS cells, only microglia express CX$_3$CR1 *in vivo*[80]. In the context of differing forms of CNS challenge, lack of CX$_3$CR1 might also have a protective role[81,82] or no role[83]. Therefore, the most reliable summary of these results is that CX$_3$CR1 signalling often modulates microglial-cell responses to CNS injury.

In addition to their inhibitory inputs from neurons, microglia differ from all other mononuclear phagocytes in that they reside, after the first few weeks of life, behind the blood–brain barrier, shielded from plasma proteins[84]. CNS trauma, infection, inflammatory and neoplastic diseases are all accompanied by a loss of blood–brain barrier integrity, and it has been considered axiomatic that exposure to plasma proteins activates microglia. This concept was rendered more specific and able to be molecularly dissected with the demonstration that plasma fibrinogen[85] is a key ingredient in the stimulation of microglia by proteins that extravasate across the disrupted blood–brain barrier. Fibrinogen, which binds to platelet integrins, also engages MAC1 (also known as CD11b–CD18 and α$_M$β$_2$-integrin), an integrin found on activated microglia.

When the domain of fibrinogen that interacts with MAC1 was identified, it was found that the interaction between the fibrinogen and microglial MAC1 can be blocked without suppressing fibrinogen's function in haemostasis. Microglia stimulated by fibrinogen had an increased phagocytic capacity, which was mediated through AKT-dependent and Rho-dependent pathways[86].

Microglia and adaptive immunity

So far, we have focused on the innate inflammatory responses of microglia. The adaptive arm of the immune response involves interactions between mononuclear phagocytes and lymphocytes and is responsible for immunological memory. The CNS parenchyma is an immunoprivileged site[14] because the only mono-nuclear phagocytes that are present — microglia — cannot assume the functions of dendritic cells, which ingest antigen, exit the tissue and enter the draining lymph nodes, where they stimulate naive T cells. Immune responses do, however, occur in the CNS, raising questions about the roles of microglia in processes such as host defence, tumour rejection and immunopathology (Table 1), as typified in individuals with multiple sclerosis.

The CNS parenchyma is inimical to T cells, which do not survive there unless restimulated by the antigen that they recognized initially in the peripheral lymphoid organs[87]. Furthermore, mononuclear phagocytes do not efficiently invade the parenchyma without the cytokine signals secreted by T cells, following this type of restimulation[88]. The restimulation of T cells by mononuclear phagocytes of the CNS has been elegantly dissected *in vivo* using genetic and imaging tools[89–94] and *in vitro* using co-culture systems[95]. The import of these accumulated observations is that the non-parenchymal mononuclear phagocytes of the CNS — those found in the choroid plexus, meninges and perivascular spaces — are responsible for cell restimulation. Despite their competence as antigen-presenting cells and their expression of dendritic-cell markers, CNS myeloid cells do not seem to have the capacity to leave the CNS and transport antigen to the lymph nodes, where immune recognition by unprimed T cells could be initiated — and this is the case even in conditions of ongoing, severe local inflammation[96]. Involvement of the parenchymal microglia seems limited to an undefined role in worsening disease severity, as has been shown in animals with experimental autoimmune encephalomyelitis, a model of immune-mediated brain inflammation[97].

Perspective

Microglia are unique mononuclear phagocytes, which are localized in the unique environment of the brain parenchyma. Understanding the properties of these cells is a challenge to biomedical scientists, who must devise strategies to study the cells *in vivo* and cope with the absence of markers that unequivocally distinguish them from haematogenous, infiltrating macrophages, which can have identical morphologies[53]. Therefore, researchers lack knowledge about the triggers and suppressors of the phenotype(s) of activated microglia. There have been recent gains in our understanding of how and when microglia enter the CNS parenchyma. A crucial line of further research asks whether and how the responses of infiltrating and resident myeloid cells might differ upon activation. Solving these mysteries will require a deeper overall understanding of how microglia interact with their environment (both within the healthy CNS parenchyma and as altered by neurodegeneration, trauma, ischaemia, infection or infiltrating immune cells). Furthermore, given the intricate anatomical specialization of the CNS, researchers will also need more profound insights into the ways in which local microenvironmental cues regulate microglial-cell behaviour.

The rewards for solving some of these conundrums — for example, a means to ameliorate neurodegeneration, to accelerate neural repair or to even deliver genetic or protein cargo to the parenchyma — could eclipse the effort expended. We await the day when researchers' capacity to modulate microglial-cell behaviour is equivalent to even a tiny fraction of that achieved effortlessly by the intact CNS. ∎

1. Mazzarello, P. The impossible interview with the man of the hidden biological structures. *J. Hist. Neurosci.* **15**, 318–325 (2006).
2. Golgi, C. On the structure of nerve cells. *Boll. Soc. Chir. Med. Pavia* **13**, 3–16 (1898); transl. Geller Lipsky, N. *J. Microsc.* **155**, 3–7 (1989).
3. Lopez-Munoz, F., Boya, J. & Alamo, C. Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. *Brain Res. Bull.* **70**, 391–405 (2006).
4. Gill, A. S. & Binder, D. K. Wilder Penfield, Pío del Río-Hortega, and the discovery of oligodendroglia. *Neurosurgery* **60**, 940–948 (2007).
5. del Río-Hortega, P. in *Cytology and Cellular Pathology of the Nervous System* (ed. Penfield, W.) 481–534 (Hoeber, 1932).
6. Nissl, F. Ueber einige Beziehungen zwischen Nervenzellerkrankungen und gliosen Erscheinungen bei verscheidenen Pschosen. *Arch. Psychiatr.* **32**, 1–21 (1899).
7. Geissmann, F., Gordon, S., Hume, D. A., Mowat, A. M. & Randolph, G. J. Unravelling mononuclear phagocyte heterogeneity. *Nature Rev. Immunol.* **10**, 453–460 (2010).
   Few review articles are as informative as this lively discourse among experts.
8. Chan, W. Y., Kohsaka, S. & Rezaie, P. The origin and cell lineage of microglia: new concepts. *Brain Res. Rev.* **53**, 344–354 (2007).
9. Galea, I., Bechmann, I. & Perry, V. H. What is immune privilege (not)? *Trends Immunol.* **28**, 12–18 (2007).
10. Fedoroff, S., Zhai, R. & Novak, J. P. Microglia and astroglia have a common progenitor cell. *J. Neurosci. Res.* **50**, 477–486 (1997).
11. Fedoroff, S. & Hao, C. Origin of microglia and their regulation by astroglia. *Adv. Exp. Med. Biol.* **296**, 135–142 (1991).
12. McKercher, S. R. *et al.* Targeted disruption of the *PU.1* gene results in multiple hematopoietic abnormalities. *EMBO J.* **15**, 5647–5658 (1996).
    This is regarded as the paper that established microglia as haematopoietic cells.
13. Lichanska, A. M. *et al.* Differentiation of the mononuclear phagocyte system during mouse embryogenesis: the role of transcription factor PU.1. *Blood* **94**, 127–138 (1999).
14. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macropages. *Science* doi:10.1126/science.1194637 (2010).
15. De Groot, C. J., Huppes, W., Sminia, T., Kraal, G. & Dijkstra, C. D. Determination of the origin and nature of brain macrophages and microglial cells in mouse central nervous system, using non-radioactive *in situ* hybridization and immunoperoxidase techniques. *Glia* **6**, 301–309 (1992).
16. Ransohoff, R. M. & Perry, V. H. Microglial physiology: unique stimuli, specialized responses. *Annu. Rev. Immunol.* **27**, 119–145 (2009).
17. Auffray, C., Sieweke, M. H. & Geissmann, F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu. Rev. Immunol.* **27**, 669–692 (2009).
18. El Khoury, J. *et al.* Ccr2 deficiency impairs microglial accumulation and accelerates progression of Alzheimer-like disease. *Nature Med.* **13**, 432–438 (2007).
    This study provided a potent demonstration of the differential capacities of microglia and monocyte-derived macrophages for responding to amyloid-β deposition.
19. El Khoury, J. & Luster, A. D. Mechanisms of microglia accumulation in Alzheimer's disease: therapeutic implications. *Trends Pharmacol. Sci.* **29**, 626–632 (2008).
20. Zhu, B. *et al.* CD11b⁺Ly-6Chi suppressive monocytes in experimental autoimmune encephalomyelitis. *J. Immunol.* **179**, 5228–5237 (2007).
21. Yong, V. W. & Rivest, S. Taking advantage of the systemic immune system to cure brain diseases. *Neuron* **64**, 55–60 (2009).
22. Tambuyzer, B. R., Ponsaerts, P. & Nouwen, E. J. Microglia: gatekeepers of central nervous system immunology. *J. Leukoc. Biol.* **85**, 352–370 (2009).
23. Sminia, T., De Groot, C. J., Dijkstra, C. D., Koetsier, J. C. & Polman, C. H. Macrophages in the central nervous system of the rat. *Immunobiology* **174**, 43–50 (1987).
24. Man, S., Ubogu, E. E. & Ransohoff, R. M. Inflammatory cell migration into the central nervous system: a few new twists on an old tale. *Brain Pathol.* **17**, 243–250 (2007).
25. Frohman, E. M. & Kerr, D. Is neuromyelitis optica distinct from multiple sclerosis? Something for 'lumpers' and 'splitters'. *Arch. Neurol.* **64**, 903–905 (2007).
26. Streit, W. J. Microglia and macrophages in the developing CNS. *Neurotoxicology* **22**, 619–624 (2001).
27. Parnaik, R., Raff, M. C. & Scholes, J. Differences between the clearance of apoptotic cells by professional and non-professional phagocytes. *Curr. Biol.* **10**, 857–860 (2000).
28. Caldero, J., Brunet, N., Ciutat, D., Hereu, M. & Esquerda, J. E. Development of microglia in the chick embryo spinal cord: implications in the regulation of motoneuronal survival and death. *J. Neurosci. Res.* **87**, 2447–2466 (2009).
29. Stevens, B. *et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178 (2007).
    This paper extends the developmental role of microglia to include synaptic pruning.
30. Ashwell, K. The distribution of microglia and cell death in the fetal rat forebrain. *Brain Res. Dev. Brain Res.* **58**, 1–12 (1991).
31. Ashwell, K. Microglia and cell death in the developing mouse cerebellum. *Brain Res. Dev. Brain Res.* **55**, 219–230 (1990).
32. Perry, V. H., Hume, D. A. & Gordon, S. Immunohistochemical localization of macrophages and microglia in the adult and developing mouse brain.

Neuroscience **15**, 313–326 (1985).
This pioneering study used myeloid markers to map the distribution of microglia in the rat CNS and establish their spatiotemporal relationship with apoptotic cells.

33. Wu, H. H. *et al.* Glial precursors clear sensory neuron corpses during development via Jedi-1, an engulfment receptor. *Nature Neurosci.* **12**, 1534–1541 (2009).

34. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).

35. Wakselman, S. *et al.* Developmental neuronal death in hippocampus requires the microglial CD11b integrin and DAP12 immunoreceptor. *J. Neurosci.* **28**, 8138–8143 (2008).

36. Frade, J. M. & Barde, Y. A. Microglia-derived nerve growth factor causes cell death in the developing retina. *Neuron* **20**, 35–41 (1998).
This paper showed that microglia help to regulate cell death in the CNS during development, as well as engulfing the cell corpses.

37. Bianchin, M. M. *et al.* Nasu–Hakola disease (polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy — PLOSL): a dementia associated with bone cystic lesions. From clinical to genetic and molecular aspects. *Cell. Mol. Neurobiol.* **24**, 1–24 (2004).

38. Tanaka, J. Nasu–Hakola disease: a review of its leukoencephalopathic and membranolipodystrophic features. *Neuropathology* **20**, S25–S29 (2000).

39. Klunemann, H. H. *et al.* The genetic causes of basal ganglia calcification, dementia, and bone cysts: DAP12 and TREM2. *Neurology* **64**, 1502–1507 (2005).

40. Neumann, H. & Takahashi, K. Essential role of the microglial triggering receptor expressed on myeloid cells-2 (TREM2) for central nervous tissue immune homeostasis. *J. Neuroimmunol.* **184**, 92–99 (2007).

41. Chouery, E. *et al.* Mutations in TREM2 lead to pure early-onset dementia without bone cysts. *Hum. Mutat.* **29**, E194–E204 (2008).
This paper showed that perturbed microglial-cell physiology, without other types of cellular pathology, can cause neurodegeneration.

42. Stefano, L. *et al.* The surface-exposed chaperone, Hsp60, is an agonist of the microglial TREM2 receptor. *J. Neurochem.* **110**, 284–294 (2009).

43. Greer, J. M. & Capecchi, M. R. Hoxb8 is required for normal grooming behavior in mice. *Neuron* **33**, 23–34 (2002).

44. Chen, S. K. *et al.* Hematopoietic origin of pathological grooming in *Hoxb8* mutant mice. *Cell* **141**, 775–785 (2010).

45. Hyman, S. E. A bone to pick with compulsive behavior. *Cell* **141**, 752–754 (2010).

46. Wake, H., Moorhouse, A. J., Jinno, S., Kohsaka, S. & Nabekura, J. Resting microglia directly monitor the functional state of synapses *in vivo* and determine the fate of ischemic terminals. *J. Neurosci.* **29**, 3974–3980 (2009).

47. Reed-Geaghan, E. G., Savage, J. C., Hise, A. G. & Landreth, G. E. CD14 and Toll-like receptors 2 and 4 are required for fibrillar Aβ-stimulated microglial activation. *J. Neurosci.* **29**, 11982–11992 (2009).

48. Lee, C. Y. & Landreth, G. E. The role of microglia in amyloid clearance from the AD brain. *J. Neural Transm.* **117**, 949–960 (2010).

49. Davalos, D. *et al.* ATP mediates rapid microglial response to local brain injury *in vivo. Nature Neurosci.* **8**, 752–758 (2005).

50. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting microglial cells are highly dynamic surveillants of brain parenchyma *in vivo. Science* **308**, 1314–1318 (2005).
References 49 and 50 established the concept of 'surveillant' microglia.

51. Kim, J. V. *et al.* Two-photon laser scanning microscopy imaging of intact spinal cord and cerebral cortex reveals requirement for CXCR6 and neuroinflammation in immune cell infiltration of cortical injury sites. *J. Immunol. Methods* **352**, 89–100 (2010).

52. Davalos, D. *et al.* Stable *in vivo* imaging of densely populated glia, axons and blood vessels in the mouse spinal cord using two-photon microscopy. *J. Neurosci. Methods* **169**, 1–7 (2008).

53. Flugel, A., Bradl, M., Kreutzberg, G. W. & Graeber, M. B. Transformation of donor-derived bone marrow precursors into host microglia during autoimmune CNS inflammation and during the retrograde response to axotomy. *J. Neurosci. Res.* **66**, 74–82 (2001).

54. Ransohoff, R. M. Microgliosis: the questions shape the answers. *Nature Neurosci.* **10**, 1507–1509 (2007).

55. Ajami, B., Bennett, J. L., Krieger, C., Tetzlaff, W. & Rossi, F. M. Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nature Neurosci.* **10**, 1538–1543 (2007).
This study used parabiosis to uncover the artefacts that are inherent in using radiation bone-marrow chimaerism to study microglial-cell physiology.

56. Mildner, A. *et al.* Microglia in the adult brain arise from Ly-6C^{hi}CCR2$^+$ monocytes only under defined host conditions. *Nature Neurosci.* **10**, 1544–1553 (2007).

57. Graeber, M. B. *et al.* The microglia/macrophage response in the neonatal rat facial nucleus following axotomy. *Brain Res.* **813**, 241–253 (1998).

58. Haynes, S. E. *et al.* The P2Y$_{12}$ receptor regulates microglial activation by extracellular nucleotides. *Nature Neurosci.* **9**, 1512–1519 (2006).

59. Orr, A. G., Orr, A. L., Li, X. J., Gross, R. E. & Traynelis, S. F. Adenosine A$_{2A}$ receptor mediates microglial process retraction. *Nature Neurosci.* **12**, 872–878 (2009).

60. Perry, V. H., Nicoll, J. A. & Holmes, C. Microglia in neurodegenerative disease. *Nature Rev. Neurol.* **6**, 193–201 (2010).

61. Mantovani, A., Sica, A. & Locati, M. Macrophage polarization comes of age. *Immunity* **23**, 344–346 (2005).

62. Shpargel, K. B. *et al.* Preconditioning paradigms and pathways in the brain. *Cleve. Clin. J. Med.* **75** (Suppl. 2), S77–S82 (2008).

63. Mirrione, M. M. *et al.* Microglial ablation and lipopolysaccharide preconditioning affects pilocarpine-induced seizures in mice. *Neurobiol. Dis.* **39**, 85–97 (2010).

64. Nakajima, K. & Kohsaka, S. Microglia activation and their significance in the central nervous system. *J. Biochem.* **130**, 169–175 (2001).

65. Nakamura, Y. Regulating factors for microglia activation. *Biol. Pharm. Bull.* **25**, 945–953 (2002).

66. Ponomarev, E. D., Shriver, L. P., Maresz, K. & Dittel, B. N. Microglial cell activation and proliferation precedes the onset of CNS autoimmunity. *J. Neurosci. Res.* **81**, 374–389 (2005).

67. Zielasek, J. & Hartung, H.-P. Molecular mechanisms of microglia activation. *Adv. Neuroimmunol.* **6**, 191–222 (1996).

68. Schmid, C. D. *et al.* Differential gene expression in LPS/IFNγ activated microglia and macrophages: *in vitro* versus *in vivo. J. Neurochem.* **109** (suppl. 1), 117–125 (2009).

69. Hanisch, U. K. & Kettenmann, H. Microglia: active sensor and versatile effector cells in the normal and pathologic brain. *Nature Neurosci.* **10**, 1387–1394 (2007).

70. Hoek, R. M. *et al.* Down-regulation of the macrophage lineage through interaction with OX2 (CD200). *Science* **290**, 1768–1771 (2000).

71. Barclay, A. N., Wright, G. J., Brooke, G. & Brown, M. H. CD200 and membrane protein interactions in the control of myeloid cells. *Trends Immunol.* **23**, 285–290 (2002).

72. Junker, A. *et al.* MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain* **132**, 3342–3352 (2009).

73. Cardona, A. & Ransohoff, R. M. Chemokine receptor CX$_3$CR1. *UCSD–Nature Molecule Pages* doi:10.1038/mp.a000633.01 (2009).

74. Harrison, J. K. *et al.* Role for neuronally derived fractalkine in mediating interactions between neurons and CX$_3$CR1-expressing microglia. *Proc. Natl Acad. Sci. USA* **95**, 10896–10901 (1998).

75. Ransohoff, R. M. Chemokines and chemokine receptors: standing at the crossroads of immunobiology and neurobiology. *Immunity* **31**, 711–721 (2009).

76. Hundhausen, C. *et al.* The disintegrin-like metalloproteinase ADAM10 is involved in constitutive cleavage of CX$_3$CL1 (fractalkine) and regulates CX$_3$CL1-mediated cell–cell adhesion. *Blood* **102**, 1186–1195 (2003).

77. Gahmberg, C. G., Tian, L., Ning, L. & Nyman-Huttunen, H. ICAM-5 — a novel two-facetted adhesion molecule in the mammalian brain. *Immunol. Lett.* **117**, 131–135 (2008).

78. Farber, K. & Kettenmann, H. Purinergic signaling and microglia. *Pflügers Arch.* **452**, 615–621 (2006).

79. Farber, K. & Kettenmann, H. Physiology of microglial cells. *Brain Res. Brain Res. Rev.* **48**, 133–143 (2005).

80. Cardona, A. E. *et al.* Control of microglial neurotoxicity by the fractalkine receptor. *Nature Neurosci.* **9**, 917–924 (2006).

81. Huang, D. *et al.* The neuronal chemokine CX$_3$CL1/fractalkine selectively recruits NK cells that modify experimental autoimmune encephalomyelitis within the central nervous system. *FASEB J.* **20**, 896–905 (2006).

82. Fuhrmann, M. *et al.* Microglial *Cx3cr1* knockout prevents neuron loss in a mouse model of Alzheimer's disease. *Nature Neurosci.* **13**, 411–413 (2010).

83. Jung, S. *et al.* Analysis of fractalkine receptor CX$_3$CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell. Biol.* **20**, 4106–4114 (2000).

84. Bechmann, I., Galea, I. & Perry, V. H. What is the blood–brain barrier (not)? *Trends Immunol.* **28**, 5–11 (2007).

85. Adams, R. A. *et al.* The fibrin-derived γ377–395 peptide inhibits microglia activation and suppresses relapsing paralysis in central nervous system autoimmune disease. *J. Exp. Med.* **204**, 571–582 (2007).
This study identified a major component of plasma that specifically activates microglia.

86. Ryu, J. K., Davalos, D. & Akassoglou, K. Fibrinogen signal transduction in the nervous system. *J. Thromb. Haemost.* **7** (suppl. 1), 151–154 (2009).

87. Chang, T. T. *et al.* Recovery from EAE is associated with decreased survival of encephalitogenic T cells in the CNS of B7-1/B7-2-deficient mice. *Eur. J. Immunol.* **33**, 2022–2032 (2003).

88. Kawakami, N. *et al.* The activation status of neuroantigen-specific T cells in the target organ determines the clinical outcome of autoimmune encephalomyelitis. *J. Exp. Med.* **199**, 185–197 (2004).

89. Hickey, W. F. & Kimura, H. Perivascular microglial cells of the CNS are bone marrow-derived and present antigen *in vivo. Science* **239**, 290–292 (1988).
This study showed that perivascular macrophages are crucial for restimulating antigen-specific T cells in the CNS.

90. Becher, B., Bechmann, I. & Greter, M. Antigen presentation in autoimmunity and CNS inflammation: how T lymphocytes recognize the brain. *J. Mol. Med.* **84**, 532–543 (2006).

91. McMahon, E. J., Bailey, S. L. & Miller, S. D. CNS dendritic cells: critical participants in CNS inflammation? *Neurochem. Int.* **49**, 195–203 (2006).

92. Bailey, S. L., Carpentier, P. A., McMahon, E. J., Begolka, W. S. & Miller, S. D. Innate and adaptive immune responses of the central nervous system. *Crit. Rev. Immunol.* **26**, 149–188 (2006).

93. Kivisakk, P. *et al.* Localizing central nervous system immune surveillance: meningeal antigen-presenting cells activate T cells during experimental autoimmune encephalomyelitis. *Ann. Neurol.* **65**, 457–469 (2009).

94. Bartholomaus, I. *et al.* Effector T cell interactions with meningeal vascular structures in nascent autoimmune CNS lesions. *Nature* **462**, 94–98 (2009).

This paper vividly dissected the interactions of antigen-specific T cells with perivascular and meningeal macrophages during extravasation of the T cells into the subarachnoid space and entry to the CNS parenchyma.

95. Aloisi, F., De, S. R., Columba-Cabezas, S., Penna, G. & Adorini, L. Functional maturation of adult mouse resting microglia into an APC is promoted by granulocyte–macrophage colony-stimulating factor and interaction with T_H1 cells. *J. Immunol.* **164,** 1705–1712 (2000).
96. McMahon, E. J., Bailey, S. L., Castenada, C. V., Waldner, H. & Miller, S. D. Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. *Nature Med.* **11,** 335–339 (2005).
97. Heppner, F. L., Greter, M., Marino, D., Falsig, J. & Raivich, G. Experimental autoimmune encephalomyelitis repressed by microglial paralysis. *Nature Med.* **11,** 146–152 (2005).
98. Kaufmann, M. H. *The Atlas of Mouse Development* (Elsevier, 1992).
99. Chitnis, T. *et al.* Elevated neuronal expression of CD200 protects *Wld^s* mice from inflammation-mediated neurodegeneration. *Am. J. Pathol.* **170,** 1695–1712 (2007).
100. Piccio, L. *et al.* Blockade of TREM-2 exacerbates experimental autoimmune encephalomyelitis. *Eur. J. Immunol.* **37,** 1290–1301 (2007).
101. Takahashi, K., Rochford, C. D. & Neumann, H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J. Exp. Med.* **201,** 647–657 (2005).
102. Mott, R. T. *et al.* Neuronal expression of CD22: novel mechanism for inhibiting microglial proinflammatory cytokine production. *Glia* **46,** 369–379 (2004).
103. Berangere, R. D. & Przedborski, S. Fractalkine: moving from chemotaxis to neuroprotection. *Nature Neurosci.* **9,** 859–861 (2006).
104. Vitkovic, L., Maeda, S. & Sternberg, E. Anti-inflammatory cytokines: expression and action in the brain. *Neuroimmunomodulation* **9,** 295–312 (2001).
105. Qian, L. *et al.* Potent anti-inflammatory and neuroprotective effects of TGF-β1 are mediated through the inhibition of ERK and p47^{phox}-Ser345 phosphorylation and translocation in microglia. *J. Immunol.* **181,** 660–668 (2008).
106. Farber, K., Pannasch, U. & Kettenmann, H. Dopamine and noradrenaline control distinct functions in rodent microglial cells. *Mol. Cell. Neurosci.* **29,** 128–138 (2005).
107. Neumann, H. & Wekerle, H. Neuronal control of the immune response in the central nervous system: linking brain immunity to neurodegeneration. *J. Neuropathol. Exp. Neurol.* **57,** 1–9 (1998).

ARTICLE

Dysfunction in GABA signalling mediates autism–like stereotypies and Rett syndrome phenotypes

Hsiao-Tuan Chao[1,2], Hongmei Chen[2], Rodney C. Samaco[2], Mingshan Xue[1]†, Maria Chahrour[2], Jong Yoo[3], Jeffrey L. Neul[4,5], Shiaoching Gong[7], Hui-Chen Lu[4,5], Nathaniel Heintz[7], Marc Ekker[8], John L. R. Rubenstein[9], Jeffrey L. Noebels[1,2,3], Christian Rosenmund[1,2]† & Huda Y. Zoghbi[1,2,3,4,5,6]

Mutations in the X-linked *MECP2* gene, which encodes the transcriptional regulator methyl-CpG-binding protein 2 (MeCP2), cause Rett syndrome and several neurodevelopmental disorders including cognitive disorders, autism, juvenile-onset schizophrenia and encephalopathy with early lethality. Rett syndrome is characterized by apparently normal early development followed by regression, motor abnormalities, seizures and features of autism, especially stereotyped behaviours. The mechanisms mediating these features are poorly understood. Here we show that mice lacking *Mecp2* from GABA (γ-aminobutyric acid)-releasing neurons recapitulate numerous Rett syndrome and autistic features, including repetitive behaviours. Loss of MeCP2 from a subset of forebrain GABAergic neurons also recapitulates many features of Rett syndrome. MeCP2-deficient GABAergic neurons show reduced inhibitory quantal size, consistent with a presynaptic reduction in glutamic acid decarboxylase 1 (*Gad1*) and glutamic acid decarboxylase 2 (*Gad2*) levels, and GABA immunoreactivity. These data demonstrate that MeCP2 is critical for normal function of GABA-releasing neurons and that subtle dysfunction of GABAergic neurons contributes to numerous neuropsychiatric phenotypes.

Mutations in *MECP2* cause Rett syndrome as well as a variety of neuropsychiatric syndromes including autism, bipolar disorder with cognitive deficits and childhood-onset schizophrenia with intellectual disabilities[1–6]. Rett syndrome is typically characterized by loss of language skills and hand use, cognitive deficits, stereotyped behaviours, impaired coordination, apraxia, ataxia, seizures, respiratory dysrhythmias and sometimes premature lethality[1,6–13]. Mice lacking MeCP2 or engineered to express an allele mimicking some Rett-syndrome-causing mutations reproduce many of the phenotypes of Rett syndrome[14–16]. In search of the anatomical basis of Rett syndrome, early studies showed that deleting MeCP2 from nearly all post-mitotic neurons and glia using Cre recombinase (Cre) driven by the nestin promoter (*Nestin*–Cre) reproduce neurological phenotypes of Rett syndrome[14,15]. In contrast, subsequent neuron-specific or regional-*Mecp2*-deletion studies reproduce some aspects of Rett syndrome[17–20]. Loss of MeCP2 from dopaminergic neurons causes motor incoordination whereas loss from serotoninergic neurons leads to increased aggression[19]; loss from the amygdala impairs amygdala-dependent learning and memory[20]; and *Mecp2* deletion in hypothalamic Sim1-expressing neurons leads to alterations in feeding behaviour, aggression and stress response[18]. Postnatal loss of MeCP2 from forebrain excitatory neurons produces motor incoordination, increased anxiety-like behaviours and impaired fear conditioning and social behaviour[17]. These genetic studies indicate that discrete features of Rett syndrome are associated with dysfunction of discrete neuronal populations (Supplementary Table 1); non-neuronal cells, such as glia, may also contribute to Rett syndrome[21,22]. Importantly, deletions of MeCP2 from specific neurons frequently provide novel insight into the role of the targeted neurons in mediating certain behaviours and phenotypes. In this study we explore the consequences of loss of MeCP2 from GABAergic neurons and demonstrate that subtle perturbation of GABAergic neuronal function contributes to neuropsychiatric phenotypes.

Generation and characterization of *Viaat*–Cre mice

Our observation that cortical wild-type GABAergic neurons express ~50% more MeCP2 than non-GABAergic neurons (Fig. 1a, $P < 0.02$) indicated that MeCP2 might be particularly important to GABAergic function. To assess the role of MeCP2 in GABAergic neurons we used a bacterial artificial chromosome containing the *Viaat* (vesicular inhibitory amino acid transporter; also known as *Slc32a1*) promoter to target Cre expression in global GABAergic neurons (Supplementary Fig. 1a–c)[23]. *Viaat* encodes a transporter required for loading GABA, as well as glycine, into synaptic vesicles[24,25]. Because the bacterial artificial chromosome contained the 77.5 kb upstream and the 23 kb downstream of the *Viaat* locus that includes the truncated coding sequence for actin-related protein 5 homologue (*Actr5*), we performed quantitative real-time reverse-transcription polymerase chain reaction (RT-qPCR) and found that neither *Viaat* nor *Actr5* is overexpressed in *Viaat*–Cre mice (Supplementary Fig. 1d). Characterization of *Viaat*–Cre expression using *Rosa26R*–eYFP (where eYFP is enhanced yellow fluorescent protein) reporter mice[26] showed >98% colocalization between GABA- and eYFP-labelled cells and no colocalization with the astrocytic marker glial fibrillary acidic protein (GFAP) (Fig. 2a–i, 3a).

[1]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA. [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [3]Department of Neurology, Baylor College of Medicine, Houston, Texas 77030, USA. [4]Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Baylor College of Medicine, Houston, Texas 77030, USA. [6]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA. [7]The Rockefeller University and Howard Hughes Medical Institute, New York, New York 10021, USA. [8]Center for Advanced Research in Environmental Genomics, Department of Biology, University of Ottawa, Ontario K1N 6N5, Canada. [9]Department of Psychiatry, University of California, San Francisco, California 94158, USA. †Present addresses: Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA (M.X.); Neurocure, Neuroscience Research Center, Charite Universitaetsmedizin Berlin, 10117, Germany (C.R.).
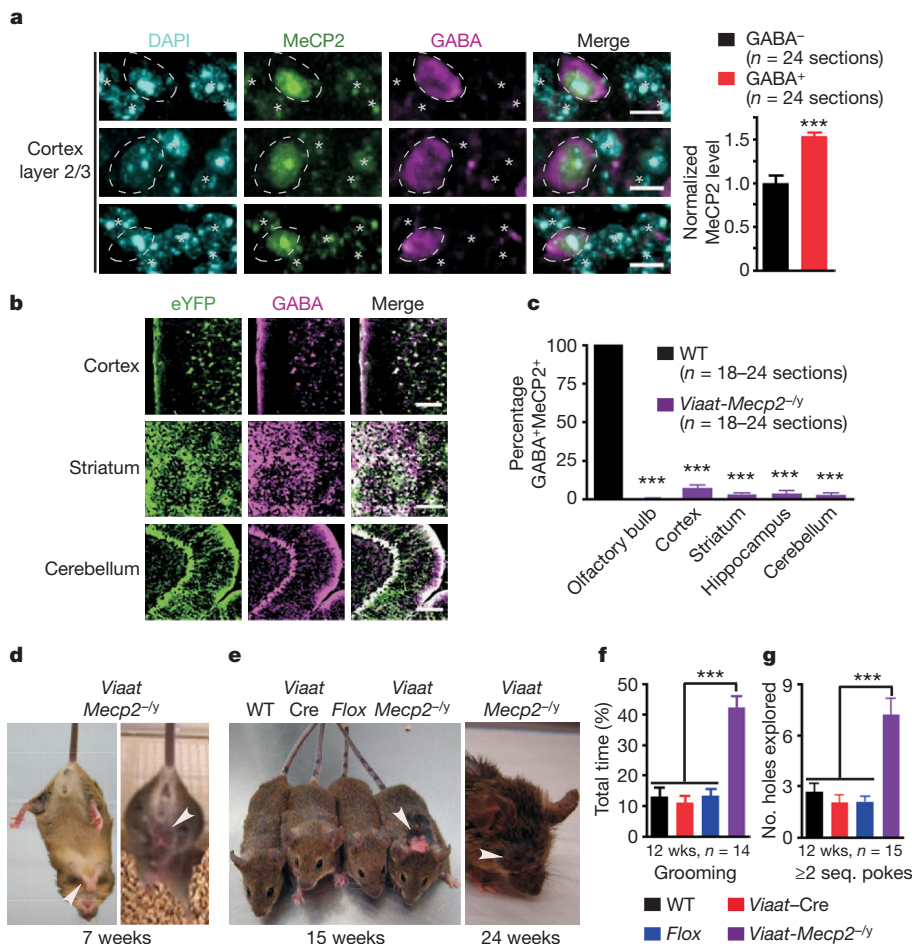
Figure 1 | *Viaat-Mecp2*$^{-/y}$ mice lose MeCP2 in GABA$^+$ neurons and develop stereotypies, self-injury and compulsive behaviour. a, Wild-type cortex layer 2/3 neurons from 17-week-old mice labelled with 4′,6-diamidino-2-phenylindole (DAPI), MeCP2 and GABA reveal 50% higher MeCP2 levels in GABA$^+$ (circled) than in GABA$^-$ cells (asterisk). Data normalized to MeCP2 level in GABA$^-$ cells; $n = 3$ mice. Scale bars, 10 μm. **b**, *Viaat–*Cre expression as assessed by *Rosa26R–*eYFP reporter and colocalization of eYFP and GABA in 14-week-old mice. Scale bars, 100 μm. **c**, More than 90% of GABA$^+$ cells in *Viaat-Mecp2*$^{-/y}$ mice lack MeCP2. Data from $n = 3$ mice per genotype. **d**, Seven-week-old *Viaat-Mecp2*$^{-/y}$ mice showing forepaw and hindlimb clasping (arrowhead). **e**, *Viaat-Mecp2*$^{-/y}$ mice showing fur loss at 15 weeks of age and self-injury, including ocular damage, at 24 weeks (arrowhead). **f**, **g**, *Viaat-Mecp2*$^{-/y}$ mice show an approximately 300% increase in grooming time (**f**) and in the number of holes explored with ≥ 2 sequential nose-pokes (seq. pokes, **g**). *Mecp2*$^{flox/y}$, *Flox*. WT, wild type; wks, weeks. Error bars are mean ± s.e.m. ***$P < 0.001$.

Viaat–Mecp2$^{-/y}$ mice show many Rett syndrome features

We generated conditional *Mecp2* deletion mice by crossing *Viaat–*Cre to *Mecp2*$^{flox/+}$ mice[14]. Male *Mecp2*-deficient mice (*Viaat-Mecp2*$^{-/y}$) lost MeCP2 from >90% of GABAergic neurons (Fig. 1c, $P < 0.0001$). We compared male *Viaat-Mecp2*$^{-/y}$ mice to male littermate controls: wild type, *cre* transgene (*Viaat-cre*), and *Mecp2*$^{flox/y}$ conditional allele (with constitutive 50% reduction in MeCP2 levels[27]).

Viaat-Mecp2$^{-/y}$ mice were indistinguishable from controls until ~5 weeks of age, when they began to exhibit repetitive behaviours such as forelimb stereotypies reminiscent of mid-line hand-wringing that characterizes Rett syndrome and hindlimb clasping (Fig. 1d, Supplementary Videos 1 and 2). These repetitive behaviours were not seen in controls (Supplementary Videos 3–5). *Viaat-Mecp2*$^{-/y}$ mice spent 300% more time grooming than wild-type mice, leading to fur loss and epidermal lesions in group- and single-housed mice (Fig. 1e, f, $P < 0.0001$). The self-injury was not due to impaired nociception (Supplementary Fig. 4a, b). Another compulsive behaviour was revealed in the holeboard assay for head-dipping stereotypy: *Viaat-Mecp2*$^{-/y}$ mice showed a greater tendency than controls to poke their nose into the same hole two or more sequential times (Fig. 1g, $P < 0.0001$).

Viaat-Mecp2$^{-/y}$ mice showed progressive motor dysfunction with more foot slips on a wire grid at 5 weeks of age (Fig. 2a, $P < 0.0001$), impaired dowel walk at 9 weeks (Fig. 2b, $P < 0.001$), and shorter latency to fall on an accelerating rotarod at 19 weeks (Supplementary Fig. 4c and Fig. 2c, $P < 0.05$). The mice also developed motor weakness with decreased latency to fall on a wire hang and reduced forelimb grip strength at 9 weeks (Fig. 2d, $P < 0.001$ and e, $P < 0.0001$). Twelve-week-old mice showed a trend towards reduced activity (Fig. 2f, $P < 0.0001$). We found

no evidence of anxiety-like phenotypes, indicating that hypoactivity did not reflect increased anxiety (Supplementary Fig. 5a–h).

A partition test and modified three-chamber assay for quantifying social interaction showed that 12–13-week-old *Viaat-Mecp2*$^{-/y}$ mice spent significantly more time showing directed interest through sniffing, pawing, or rearing near the partner mouse (Fig. 2g, $P < 0.0001$; and h, $P < 0.0001$; Supplementary Fig. 6c–f), but in a separate novel object interaction test they showed no more interest in a novel Lego inanimate object than controls (Fig. 2i). The mice were poor nest builders (Fig. 2j, $P < 0.0001$, and Supplementary Fig. 6a, b, $P < 0.0001$), which could be due to alterations in social behaviour or forepaw apraxia. As we found no signs of anxiety-like phenotypes, these alterations in social behaviour are independent of anxiety (Supplementary Fig. 5a–h).

We assessed olfactory recognition and learning and found that *Viaat-Mecp2*$^{-/y}$ mice recognized and habituated to a novel vanilla odorant, but spent more time sniffing vanilla than controls on the first test day without differences on the second test day (Supplementary Fig. 7a, b, $P < 0.01$). This increased sniffing time could be a manifestation of repetitive behaviour and may contribute to the increase in social interaction.

We examined the acoustic startle response and percentage of pre-pulse inhibition as indicators of sensorimotor arousal and gating[28]. *Viaat-Mecp2*$^{-/y}$ mice showed a significantly lower acoustic startle response to 120 dB than controls (Fig. 2k, $P < 0.001$) and significantly increased prepulse inhibition at 78 and 82 dB prepulses (Fig. 2l, $P < 0.05$), revealing that GABAergic regulation of sensorimotor gating and arousal requires MeCP2 function.

To determine whether MeCP2 deficiency in GABAergic neurons impairs hippocampal learning and memory, we evaluated *Viaat-Mecp2*$^{-/y}$ mice in the Morris water maze paradigm. Despite a similar
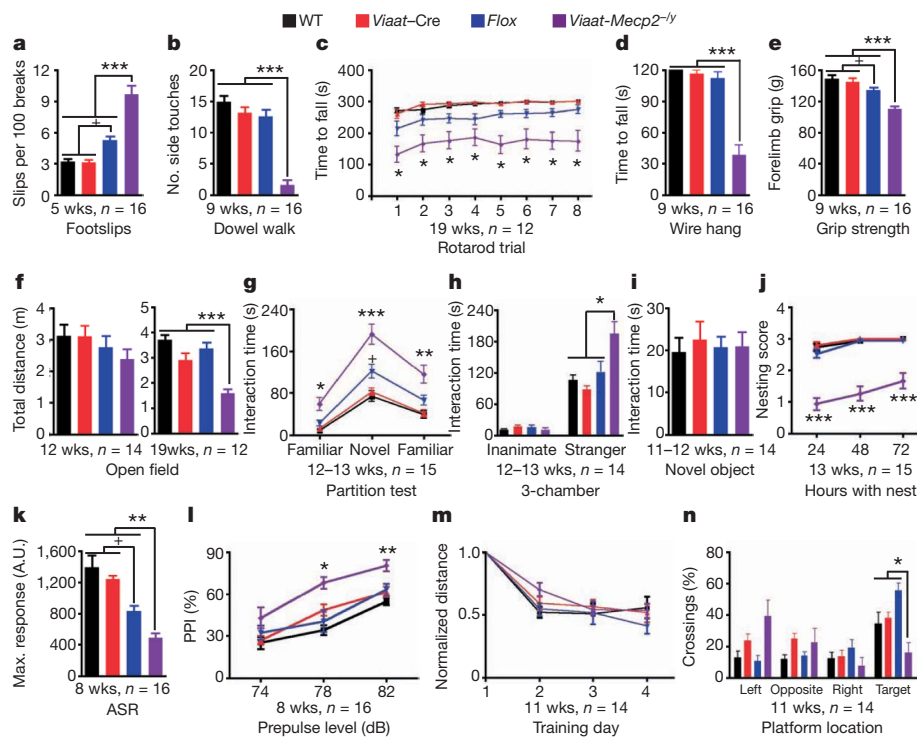
Figure 2 | MeCP2 deficiency in GABAergic neurons causes several Rett syndrome-like features. a–f, *Viaat-Mecp2*[−/y] mice show more footslips (**a**), a reduced number of side touches on a dowel (**b**), shorter latency to fall on a rotarod (**c**) and wire (**d**), reduced forelimb grip strength (**e**) and pronounced hypoactivity (**f**). **g,** *Viaat-Mecp2*[−/y] mice show intact social recognition but increased social interaction with novel and familiar partners. **h,** *Viaat-Mecp2*[−/y] mice spend 60% more time interacting with an unfamiliar mouse than compared to controls. The wire cup served as a familiar inanimate control without social valence. **i,** *Viaat-Mecp2*[−/y] mice show a similar interaction time with a novel inanimate Lego object compared to controls. **j,** *Viaat-Mecp2*[−/y] mice are poor nest builders. **k, l,** *Viaat-Mecp2*[−/y] mice have an impaired maximum acoustic startle response (ASR) to 120 dB (**k**) and increased prepulse inhibition (PPI) at 78 and 82 dB prepulses (**l**). A.U., arbitrary units. **m, n,** *Viaat-Mecp2*[−/y] mice show a similar learning rate during training (**m**) but reduced crossings over the target platform location during the probe test (**n**). Error bars are mean ± s.e.m. *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$.

rate of learning during the four training days, *Viaat-Mecp2*[−/y] mice had difficulty locating the platform during the probe trial (Fig. 2m, n, $P < 0.05$).

Premature lethality is associated with MeCP2 deficiency in all human males, in some Rett syndrome females, and in mouse models lacking MeCP2 in all neurons[1,14,15]. Roughly half of *Viaat-Mecp2*[−/y] mice died by 26 weeks of age after a period of marked weight loss (Fig. 3a and Supplementary Fig. 8). Coinciding with the weight loss, *Viaat-Mecp2*[−/y] mice developed severe respiratory dysfunction, with a 42% reduction in tidal volume, a 45% reduction in minute volume and frequent apnoeas lasting for more than 0.4 s (Fig. 3b, c, $P < 0.0001$, d, $P < 0.02$, and e, $P < 0.05$). Similar respiratory abnormalities are prominent in Rett syndrome and could contribute to early death[8,9]. We note that *Mecp2*[flox/y] males, which show some respiratory alterations[19,27], show neither premature lethality nor apnoeas; thus, constitutive partial reduction of MeCP2 levels is insufficient to shorten lifespan.

Dlx5/6-Mecp2[−/y] mice show some Rett syndrome features

Because dysfunction of forebrain structures, such as the striatum and cortex, could be responsible for the stereotypies, compulsive behaviour, motor dysfunction and social behavioural alterations seen in Rett syndrome, we used a forebrain-specific GABAergic Cre recombinase, *Dlx5/6*–Cre[29,30], to remove MeCP2 from a subset of forebrain GABAergic neurons and generate male conditional deletion mice referred to as *Dlx5/6-Mecp2*[−/y]. *Dlx5/6-Mecp2*[−/y] mice showed repetitive behaviour, impaired motor coordination, increased social interaction preference, reduced acoustic startle response and enhanced prepulse inhibition. Complete details of the characterization of the *Dlx5/6*–Cre expression pattern and *Dlx5/6-Mecp2*[−/y] phenotypes can be found in Supplementary Figs 3b, 7c, d and Supplementary Figs 9–12. In contrast to *Viaat-Mecp2*[−/y] mice, *Dlx5/6-Mecp2*[−/y] mice survive to at least 80 weeks without apparent alterations in respiratory function (Fig. 3a, b, f–h), indicating that respiratory dysfunction and

Figure 3 | Loss of MeCP2 in inhibitory GABAergic neurons compromises respiration and survival. a, *Viaat-Mecp2*[−/y] mice show premature lethality with 50% survival by 26 weeks. *Dlx5/6-Mecp2*[−/y] mice survive for at least 80 weeks. **b,** Representative plethysmography traces from 32-week-old wild-type, *Flox*, *Dlx5/6*–Cre, *Viaat*–Cre, *Dlx5/6-Mecp2*[−/y] and *Viaat-Mecp2*[−/y] mice. Only *Viaat-Mecp2*[−/y] mice show pronounced apnoeas and an abnormal respiratory pattern. **c–e,** *Viaat-Mecp2*[−/y] mice show a 42% reduction in tidal volume (**c**), a 45% reduction in minute volume (**d**) and an increased number of apnoeas longer than 0.4 s (**e**). **f–h,** *Dlx5/6-Mecp2*[−/y] mice show no alterations in tidal volume or minute volume, and no apnoeas. Error bars are mean ± s.e.m. *$P < 0.05$, ***$P < 0.001$.
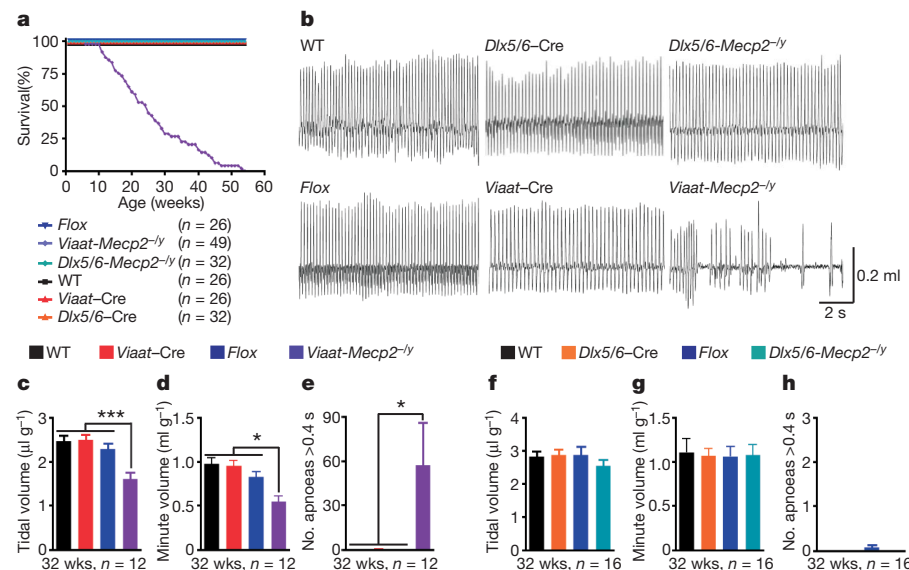
premature lethality could result either from loss of MeCP2 in hindbrain GABAergic neurons only or synergism between MeCP2 deficiency in hindbrain and forebrain GABAergic neurons.

Reduced GABA and *Gad1/2* in *Viaat-Mecp2*$^{-/y}$ mice

Given the repetitive behaviours, motor deficits, altered sensorimotor gating and arousal, and altered social behaviour observed in *Viaat-Mecp2*$^{-/y}$ and *Dlx5/6-Mecp2*$^{-/y}$ mice, we focused on cortex and striatum to determine a mechanism by which selective MeCP2 deficiency affects GABAergic function.

Immunolabelling showed that somatic GABA content in layer 2/3 cortical neurons and in striatal medium spiny neurons was reduced by 37% and 50%, respectively, in *Viaat-Mecp2*$^{-/y}$ mice (Fig. 4a, $P < 0.0005$, and b, $P < 0.0001$). Given that somatic GABA content is indicative of intracellular GABA level, which is determined by the biosynthetic activity of glutamic acid decarboxylase 67 (Gad67) and

glutamic acid decarboxylase 65 (Gad65)[31], we assessed whether mRNA levels of *Gad1* (encoding Gad67) and *Gad2* (encoding Gad65) are affected by MeCP2 deficiency in GABAergic neurons. RT–qPCR revealed that mRNA levels of *Gad1* and *Gad2* were reduced by 36% and 28%, respectively, in *Viaat-Mecp2*$^{-/y}$ cortex (Fig. 4c, $P < 0.05$) and reduced by 54% and 62%, respectively, in striatum (Fig. 4d, $P < 0.05$). To determine whether the reduction in *Gad1* and *Gad2* expression is specific to MeCP2 deletion from GABAergic neurons, we used *CamKIIα–Cre*[32] to selectively delete *Mecp2* from forebrain excitatory neurons (denoted as *CamKIIα-Mecp2*$^{-/y}$ mice). Neither *Gad1* nor *Gad2* expression is altered in *CamKIIα-Mecp2*$^{-/y}$ cortex (Fig. 4e), demonstrating the cell-autonomous specificity of reduced *Gad1* and *Gad2* expression in *Viaat-Mecp2*$^{-/y}$ mice.

Chromatin immunoprecipitation (ChIP) on wild-type brain tissue showed that MeCP2 occupies the region 1 kb upstream of the transcription start site of *Gad1* and *Gad2* (Fig. 4f), and that this
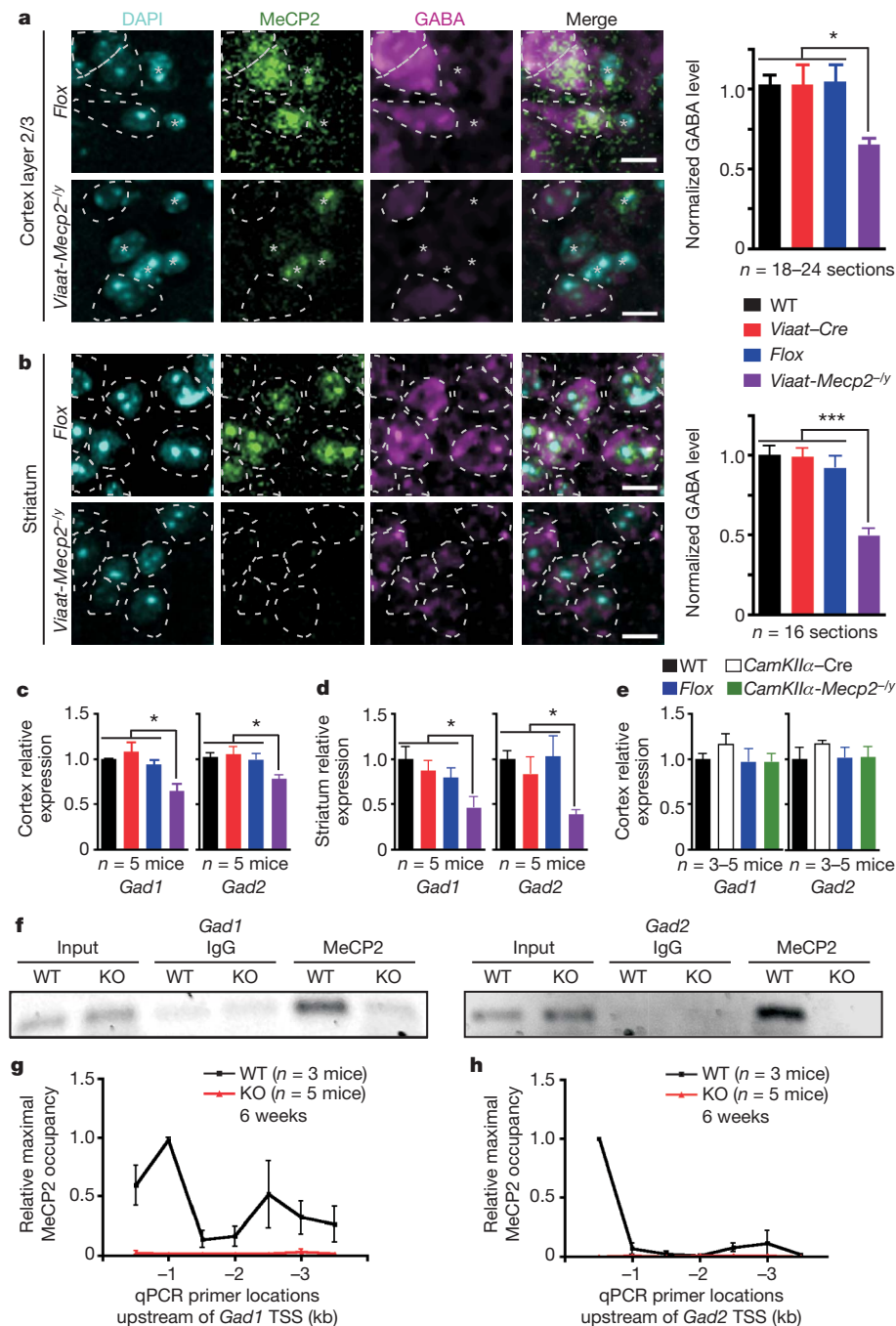
Figure 4 | MeCP2 deficiency in GABAergic neurons reduces *Gad1*, *Gad2* and GABA levels. a, b, Somatic GABA immunoreactivity in GABA$^+$ cells from 15–17-week old *Flox* and *Viaat-Mecp2*$^{-/y}$ mice labelled with DAPI, MeCP2 and GABA. GABA$^-$ cells are labelled with asterisks. Images show loss of MeCP2 and reduced GABA immunoreactivity in GABA$^+$ cells (circled) of *Viaat-Mecp2*$^{-/y}$ mice by 37% in cortical layer 2/3 neurons (**a**) and 50% in striatal neurons (**b**). Data normalized to GABA level in wild type; $n = 2$–4 mice per genotype. Scale bars, 10 μm. **c, d,** *Gad1* and *Gad2* mRNA levels are reduced in *Viaat-Mecp2*$^{-/y}$ cortex by 36% and 28%, respectively (**c**), and in *Viaat-Mecp2*$^{-/y}$ striatum by 54% and 62%, respectively (**d**). Colour key is the same as that for **a** and **b**. **e,** *Gad1* and *Gad2* mRNA levels are unaltered in *CamKIIα-Mecp2*$^{-/y}$ cortex. **f,** ChIP reveals MeCP2 occupancy of *Gad1* and *Gad2* promoters in wild type, which is absent in IgG and knockout mice. **g, h,** Mapping MeCP2 occupancy upstream of the transcription start site, after normalization with IgG, reveals increased occupancy in wild type (black line) across *Gad1* (**g**) and *Gad2* (**h**) promoters without enhanced MeCP2 binding in knockout mice (red line). KO, knockout; TSS, transcription start site. Error bars are mean ± s.e.m. *$P < 0.05$, ***$P < 0.001$.

binding is absent in constitutive null ($Mecp2^{-/y}$, or knockout)[14] and IgG controls. Using primers spanning the region from 500 bp to 3.5 kb upstream of the $Gad1$ and $Gad2$ transcription start site, we identified selective enrichment of MeCP2 occupancy upstream of the transcription start site for both genes and one intergenic region for $Gad1$ (Fig. 4g, h) demonstrating enrichment of MeCP2 at the $Gad1$ and $Gad2$ promoters. Promoter occupancy for $Gad1$ and $Gad2$ was minimal in knockout brain tissue (Fig. 4g, h), confirming the binding specificity. Reduced GABA immunoreactivity and $Gad1/2$ expression predict that MeCP2 deficiency in GABAergic neurons may reduce GABAergic synaptic quantal size.

Altered quantal size, excitability and plasticity

Recording of miniature inhibitory postsynaptic currents (mIPSCs) from layer 2/3 pyramidal neurons of the somatosensory (S1) cortex in acute brain slices showed that the amplitude and charge of mIPSCs were reduced in $Viaat$-$Mecp2^{-/y}$ mice, with no alterations in frequency (Fig. 5a, b, $P < 0.0001$). However, the miniature excitatory postsynaptic currents (mEPSCs) quantal size and frequency were

similar to controls (Fig. 5c, d), indicating that MeCP2 deficiency in GABAergic neurons does not have a non-cell-autonomous impact on quantal release from glutamatergic neurons.

To determine whether the reduction in mIPSC quantal size is due to pre- or postsynaptic alterations, we used cultured neurons for pharmacological manipulation. We confirmed that MeCP2-deficient autaptic striatal GABAergic neurons showed reduced mIPSC amplitude and charge, with no alterations in frequency (Fig. 5e, $P < 0.02$, f). Next we examined responsiveness of GABA$_A$ receptors by applying a 2-s pulse of 5 µM GABA and found no differences in the magnitude of the response between wild-type and knockout neurons (Fig. 5g), indicating that MeCP2 deficiency does not significantly alter the postsynaptic response to GABA. MeCP2 deficiency also does not significantly alter the paired pulse ratio, indicating that presynaptic release probability is similar between wild-type and knockout neurons (Fig. 5h). These results show that MeCP2 deficiency in GABAergic neurons leads to functional presynaptic reduction in GABA release and are consistent with reduced $Gad1/2$ levels and GABA immunoreactivity. Altogether, these data indicate that loss of MeCP2 in
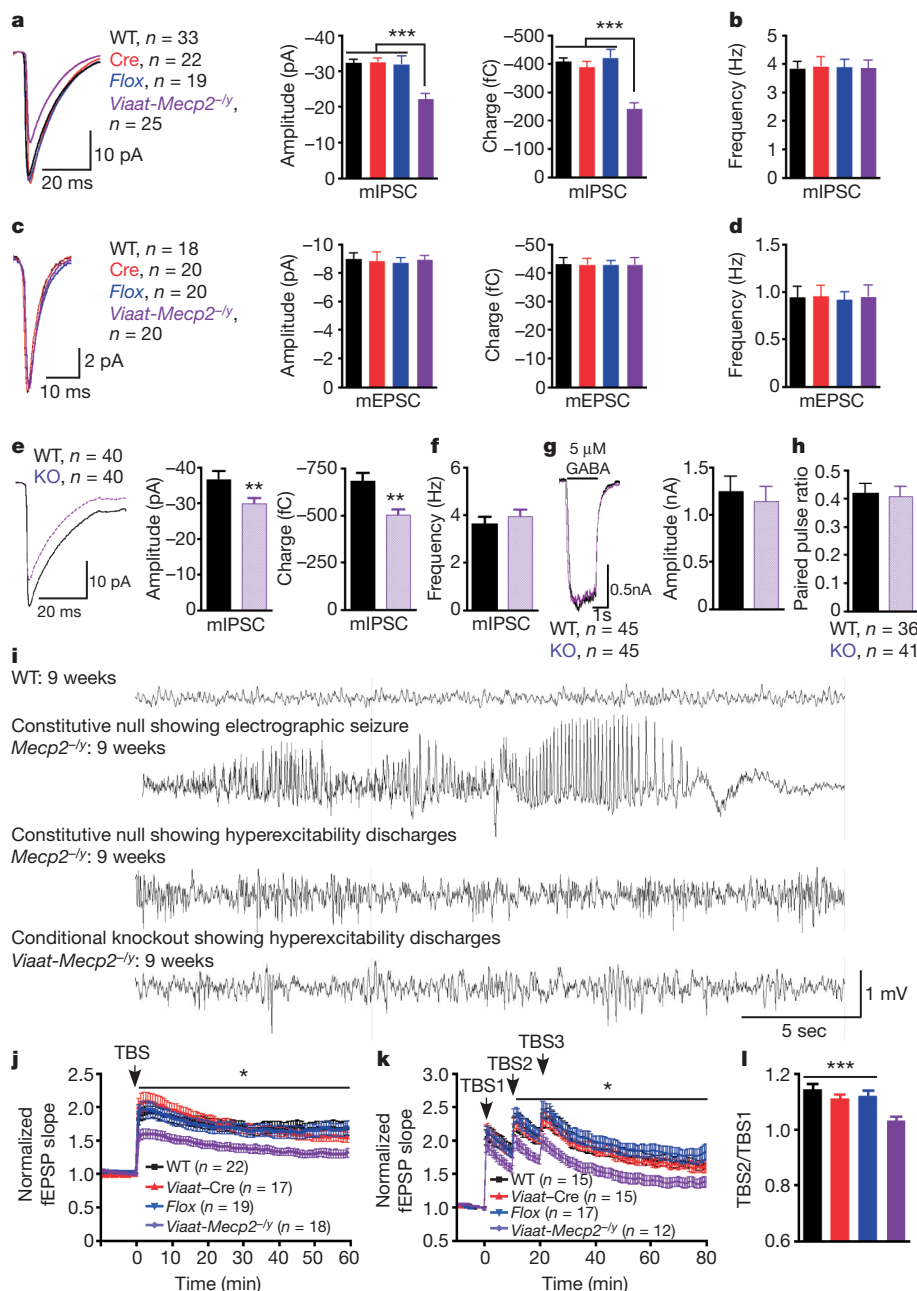
Figure 5 | MeCP2 deficiency in GABAergic neurons results in reduced mIPSC quantal size in cortical layer 2/3 and striatal neurons, EEG hyperexcitability and impaired hippocampal LTP. a–d, Data from three mice per genotype and the number of neurons recorded is shown. **a**, **b**, mIPSC amplitude and charge are reduced in $Viaat$-$Mecp2^{-/y}$ cortical slices. Average traces for each genotype are overlaid and graphs show mIPSC amplitude and charge (**a**), and frequency (**b**). **c**, **d**, mEPSC amplitude and charge are unaltered in $Viaat$-$Mecp2^{-/y}$ cortical slices. Average traces for each genotype are overlaid and graphs show mEPSC amplitude and charge (**c**), and frequency (**d**). **e–h**, Data from two independent autaptic striatal cultures and the number of neurons recorded are shown. Average traces for each genotype are overlaid and bar graphs show that knockout neurons have reduced mIPSC amplitude and charge (**e**), but no differences in frequency (**f**). **g**, 5 µM GABA evokes similarly sized responses from wild-type and knockout neurons. **h**, Paired pulse ratio is similar between wild-type and knockout neurons. **i**, EEG recordings from constitutive null and $Viaat$-$Mecp2^{-/y}$ mice compared to wild type. Constitutive null $Mecp2^{-/y}$ mice ($n = 4$) occasionally develop electrographic seizures, but predominantly show hyperexcitability discharges. $Viaat$-$Mecp2^{-/y}$ mice ($n = 7$) frequently show hyperexcitability discharges, but do not show electrographic seizures. **j–l**, Acute hippocampal slices from 11–13-week old mice, six mice per genotype, reveal reduced magnitude of theta burst stimulation induced LTP (**j**) and saturating LTP leads to no further increases in synaptic potentiation (**k**) in $Viaat$-$Mecp2^{-/y}$ slices. Number of slices recorded is shown in the figure. fEPSP, field excitatory postsynaptic potential; TBS, theta-burst stimulation. **l**, There is a significant increase in potentiation with the second TBS in controls but not in $Viaat$-$Mecp2^{-/y}$ slices. Number of slices recorded shown in **k**. Error bars are mean ± s.e.m. **$P < 0.01$, ***$P < 0.001$.

GABAergic neurons compromises intracellular GABA levels via a cell-autonomous effect on *Gad1/2* mRNA levels.

The reduction in GABA content and inhibitory neurotransmission could cause hyperexcitable network activity. Electroencephalographic (EEG) recordings revealed that constitutive null mice suffer both electrographic seizure and non-seizure hyperexcitability discharges (Fig. 5i). *Viaat-Mecp2$^{-/y}$* mice showed frequent hyperexcitability discharges (Fig. 5i), but no electrographic seizures despite prolonged continuous monitoring. Remarkably, EEG recordings from 39-week-old *Dlx5/6-Mecp2$^{-/y}$* mice showed neither electrographic seizures nor hyperexcitability discharges (Supplementary Fig. 10l), indicating that loss of MeCP2 in a subset of GABAergic neurons is insufficient to cause hyperexcitability.

To determine whether impairments in synaptic plasticity can result from loss of MeCP2 in GABAergic neurons, we examined Schaffer collateral synapses in acute hippocampal slices from *Viaat-Mecp2$^{-/y}$* mice and controls. We observed no significant differences in input-output curves or paired-pulse ratios (Supplementary Fig. 13a, b), but we did observe impaired long-term potentiation (LTP) induced by theta-burst stimulation of Schaffer collateral synapses in *Viaat-Mecp2$^{-/y}$* mice (Fig. 5j, $P < 0.05$). To determine whether additional stimulation could increase LTP in *Viaat-Mecp2$^{-/y}$* slices, we used a saturating LTP paradigm with three separate theta-burst stimulations. In contrast with control slices, *Viaat-Mecp2$^{-/y}$* slices resisted further increases in potentiation (Fig. 5k, l, $P < 0.05$), possibly because the basal state of the circuit in *Viaat-Mecp2$^{-/y}$* slices is closer to maximal potentiation and therefore cannot respond with further potentiation on subsequent stimulations. These results also indicate that cellular disruptions or circuit alterations in addition to altered GABAergic transmission may contribute to impaired LTP.

Discussion

Previous electrophysiological studies using mice constitutively lacking MeCP2 showed a reduction in excitatory synaptic strength and glutamatergic synapse numbers in both isolated autaptic neurons and slices[33,34]. Furthermore, some studies indicated that the strength of inhibition in the cortex, hippocampus and brainstem is also altered upon constitutive loss of MeCP2[35–37]. Whether these alterations in GABAergic inhibition were cell autonomous or had specific behavioural consequences remained unknown. Here we demonstrate that GABAergic dysfunction is a critical mediator of Rett syndrome phenotypes. Mice with MeCP2 deficiency in GABAergic neurons initially show normal behaviour, then develop forepaw stereotyped movements, compulsive grooming, increased sociability (similar to Rett syndrome stage III or IV)[12], impaired motor coordination, learning/memory deficits, abnormal EEG hyperexcitability, severe respiratory dysrhythmias and premature lethality. Sensorimotor gating and arousal are also altered. It is noteworthy that MeCP2 deficiency in a subset of forebrain GABAergic neurons is sufficient to cause some repetitive behaviours, impaired motor coordination, increased social interest, and altered sensorimotor gating and arousal. The phenotypic similarity between *Viaat-Mecp2$^{-/y}$* mice and Rett syndrome strongly indicates that disrupting MeCP2 function in GABAergic neurons alone perturbs the neuronal network in a way similar to that which occurs in Rett syndrome.

It is intriguing that subtle disruption of GABAergic neuronal function by a 30–40% reduction in GABA neurotransmitter release, as assayed neurophysiologically, leads to significant functional abnormalities and neuropsychiatric features. Previous studies indicated a role for GABAergic dysfunction in isolated behaviours, such as altered learning and memory[38,39] and impaired social behaviour[40] (but see also ref. 41). To our knowledge, loss of MeCP2 in GABAergic neurons, both globally and in a subset, is unique in that it reveals a multitude of neuropsychiatric phenotypes encompassing social behaviour, learning/memory, motor function, stereotyped behaviours and sensorimotor gating.

The discovery that MeCP2 deficiency in GABAergic neurons alters *Gad1/2* expression and results in a change in neuronal GABA content with corresponding alterations in synaptic physiology provides critical insight into how MeCP2 supports GABAergic function. Although MeCP2 binds broadly throughout the genome with selectivity to methylated sites[42], there are sites where binding is increased relative to neighbouring regions, including alternative promoters at the *Bdnf* locus[42–45] and, as shown in this study, for *Gad1/2*. Exactly how the broad and region-specific binding of MeCP2 affects specific gene expression remains to be elucidated but certainly the reduction in *Bdnf* messenger RNA levels (in *Mecp2$^{-/y}$* mice) and *Gad1/2* (in this study) demonstrate that such binding is of consequence to gene expression. It is worth noting in this context that some post-mortem studies have observed reduced *Gad1/2* expression in schizophrenia, bipolar disorder, or autism patients[46–50]; three disorders found, in rare instances, to be caused by *MECP2* mutations[2–5]. It may be that GABAergic dysregulation, from any of a variety of causes, is central to the pathogenesis of these neuropsychiatric disorders. Our study supports a critical role of the GABAergic system in modulating neuropsychiatric behaviours and indicates that manipulations of this system may be a promising avenue for therapeutic interventions.

METHODS SUMMARY

The mice used in all behavioural, molecular and electrophysiological studies were still active, still gaining weight, had no signs of seizures (before, during, or after behavioural testing), did not display any staring-freezing episodes and had no signs of delay in performing the behavioural tasks. All data are presented as mean ± standard error of mean, with n and ages as shown in figures or stated in text. Full statistical tests and values for behavioural studies are presented in Supplementary Tables 2 and 3. *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$.

1. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56**, 422–437 (2007).
2. Lam, C. W. *et al.* Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. *J. Med. Genet.* **37**, e41 (2000).
3. Klauck, S. M. *et al.* A mutation hot spot for nonspecific X-linked mental retardation in the *MECP2* gene causes the PPM-X syndrome. *Am. J. Hum. Genet.* **70**, 1034–1037 (2002).
4. Cohen, D. *et al.* MECP2 mutation in a boy with language disorder and schizophrenia. *Am. J. Psychiatry* **159**, 148–149 (2002).
5. Carney, R. M. *et al.* Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr. Neurol.* **28**, 205–211 (2003).
6. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185–188 (1999).
7. Hagberg, B., Aicardi, J., Dias, K. & Ramos, O. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* **14**, 471–479 (1983).
8. Weese-Mayer, D. E. *et al.* Autonomic nervous system dysregulation: breathing and heart rate perturbation during wakefulness in young girls with Rett syndrome. *Pediatr. Res.* **60**, 443–449 (2006).
9. Weese-Mayer, D. E. *et al.* Autonomic dysregulation in young girls with Rett syndrome during nighttime in-home recordings. *Pediatr. Pulmonol.* **43**, 1045–1060 (2008).
10. Deidrick, K. M., Percy, A. K., Schanen, N. C., Mamounas, L. & Maria, B. L. Rett syndrome: pathogenesis, diagnosis, strategies, therapies, and future research directions. *J. Child Neurol.* **20**, 708–717 (2005).
11. Jedele, K. B. The overlapping spectrum of Rett and Angelman syndromes: a clinical review. *Semin. Pediatr. Neurol.* **14**, 108–117 (2007).
12. Hagberg, B. Clinical manifestations and stages of Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 61–65 (2002).
13. Neul, J. L. *et al.* Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* **70**, 1313–1321 (2008).
14. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genet.* **27**, 322–326 (2001).
15. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nature Genet.* **27**, 327–331 (2001).
16. Shahbazian, M. *et al.* Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35**, 243–254 (2002).
17. Gemelli, T. *et al.* Postnatal loss of methyl-CpG binding protein 2 in the forebrain is sufficient to mediate behavioral aspects of Rett syndrome in mice. *Biol. Psychiatry* **59**, 468–476 (2006).

18. Fyffe, S. L. *et al.* Deletion of *Mecp2* in Sim1-expressing neurons reveals a critical role for MeCP2 in feeding behavior, aggression, and the response to stress. *Neuron* **59,** 947–958 (2008).
19. Samaco, R. C. *et al.* Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis and specific behavioral abnormalities. *Proc. Natl Acad. Sci. USA* **106,** 21966–21971 (2009).
20. Adachi, M., Autry, A. E., Covington, H. E. III & Monteggia, L. M. MeCP2-mediated transcription repression in the basolateral amygdala may underlie heightened anxiety in a mouse model of Rett syndrome. *J. Neurosci.* **29,** 4218–4227 (2009).
21. Ballas, N., Lioy, D. T., Grunseich, C. & Mandel, G. Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nature Neurosci.* **12,** 311–317 (2009).
22. Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J. M. & Jin, L. W. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci.* **29,** 5051–5061 (2009).
23. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27,** 9817–9823 (2007).
24. Chaudhry, F. A. *et al.* The vesicular GABA transporter, VGAT, localizes to synaptic vesicles in sets of glycinergic as well as GABAergic neurons. *J. Neurosci.* **18,** 9733–9750 (1998).
25. Wojcik, S. M. *et al.* A shared vesicular carrier allows synaptic corelease of GABA and glycine. *Neuron* **50,** 575–587 (2006).
26. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1,** 4 (2001).
27. Samaco, R. C. *et al.* A partial loss of function allele of methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. *Hum. Mol. Genet.* **17,** 1718–1727 (2008).
28. Swerdlow, N. R., Geyer, M. A. & Braff, D. L. Neural circuit regulation of prepulse inhibition of startle in the rat: current knowledge and future challenges. *Psychopharmacology (Berl.)* **156,** 194–215 (2001).
29. Monory, K. *et al.* The endocannabinoid system controls key epileptogenic circuits in the hippocampus. *Neuron* **51,** 455–466 (2006).
30. Kohwi, M. *et al.* A subpopulation of olfactory bulb GABAergic interneurons is derived from *Emx1*- and *Dlx5/6*-expressing progenitors. *J. Neurosci.* **27,** 6878–6891 (2007).
31. Martin, D. L. & Rimvall, K. Regulation of γ-aminobutyric acid synthesis in the brain. *J. Neurochem.* **60,** 395–407 (1993).
32. Tsien, J. Z. *et al.* Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **87,** 1317–1326 (1996).
33. Chao, H. T., Zoghbi, H. Y. & Rosenmund, C. MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* **56,** 58–65 (2007).
34. Dani, V. S. & Nelson, S. B. Intact long-term potentiation but reduced connectivity between neocortical layer 5 pyramidal neurons in a mouse model of Rett Syndrome. *J. Neurosci.* **29,** 11263–11270 (2009).
35. Dani, V. S. *et al.* Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA* **102,** 12560–12565 (2005).
36. Medrihan, L. *et al.* Early defects of GABAergic synapses in the brain stem of a MeCP2 mouse model of Rett syndrome. *J. Neurophysiol.* **99,** 112–121 (2008).
37. Zhang, L., He, J., Jugloff, D. G. & Eubanks, J. H. The MeCP2-null mouse hippocampus displays altered basal inhibitory rhythms and is prone to hyperexcitability. *Hippocampus* **18,** 294–309 (2008).
38. Cui, Y. *et al.* Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* **135,** 549–560 (2008).
39. Fernandez, F. *et al.* Pharmacotherapy for cognitive impairment in a mouse model of Down syndrome. *Nature Neurosci.* **10,** 411–413 (2007).
40. Tabuchi, K. *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318,** 71–76 (2007).
41. Chadman, K. K. *et al.* Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. *Autism Res.* **1,** 147–158 (2008).
42. Skene, P. J. *et al.* Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37,** 457–468 (2010).
43. Yasui, D. H. *et al.* Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl Acad. Sci. USA* **104,** 19416–19421 (2007).
44. Chen, W. G. *et al.* Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* **302,** 885–889 (2003).
45. Martinowich, K. *et al.* DNA methylation-related chromatin remodeling in activity-dependent *Bdnf* gene regulation. *Science* **302,** 890–893 (2003).
46. Akbarian, S. *et al.* Gene expression for glutamic acid decarboxylase is reduced without loss of neurons in prefrontal cortex of schizophrenics. *Arch. Gen. Psychiatry* **52,** 258–266 (1995).
47. Fatemi, S. H. *et al.* Glutamic acid decarboxylase 65 and 67 kDa proteins are reduced in autistic parietal and cerebellar cortices. *Biol. Psychiatry* **52,** 805–810 (2002).
48. Addington, A. M. *et al. GAD1* (2q31.1), which encodes glutamic acid decarboxylase (GAD₆₇), is associated with childhood-onset schizophrenia and cortical gray matter volume loss. *Mol. Psychiatry* **10,** 581–588 (2005).
49. Lundorf, M. D. *et al.* Mutational screening and association study of glutamate decarboxylase 1 as a candidate susceptibility gene for bipolar affective disorder and schizophrenia. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **135B,** 94–101 (2005).
50. Fatemi, S. H., Stary, J. M., Earle, J. A., Araghi-Niknam, M. & Eagan, E. GABAergic dysfunction in schizophrenia and mood disorders as reflected by decreased levels of glutamic acid decarboxylase 65 and 67 kDa and Reelin proteins in cerebellum. *Schizophr. Res.* **72,** 109–122 (2005).

Author Contributions H.-T.C. and H.Y.Z. conceived the study. H.-T.C., M.X., C.R. and H.Y.Z. designed experiments with input from H.C., R.C.S., J. L. Neul, H.-C.L. and J. L. Noebels. H.-T.C., H.C., R.C.S., M.X., M.C., J.Y. and J. L. Neul performed experiments. H.-T.C., H.C., M.X., J.Y. and J. L. Neul analysed data; H.-T.C., M.X., C.R. and H.Y.Z. interpreted data with input from H.C., R.C.S., J.Y., J. L. Neul, H.-C.L. and J. L. Noebels. S.G. and N.H. provided reagents for generation of *Viaat*–Cre; J.L.R.R. and M.E. provided *Dlx5/6*–Cre mice. H.-T.C., M.X. and H.Y.Z. wrote the manuscript and H.C., R.C.S., M.C., J.L. Neul, S.G., J.L.R.R, J. L. Noebels and C.R. provided input.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.R. (christian.rosenmund@charite.de) or H.Y.Z. (hzoghbi@bcm.edu).

ARTICLE

Genetic dissection of an amygdala microcircuit that gates conditioned fear

Wulf Haubensak[1], Prabhat S. Kunwar[1]*, Haijiang Cai[1]*, Stephane Ciocchi[3]*, Nicholas R. Wall[4], Ravikumar Ponnusamy[5], Jonathan Biag[6], Hong-Wei Dong[6], Karl Deisseroth[7], Edward M. Callaway[4], Michael S. Fanselow[5], Andreas Lüthi[3] & David J. Anderson[1,2]

The role of different amygdala nuclei (neuroanatomical subdivisions) in processing Pavlovian conditioned fear has been studied extensively, but the function of the heterogeneous neuronal subtypes within these nuclei remains poorly understood. Here we use molecular genetic approaches to map the functional connectivity of a subpopulation of GABA-containing neurons, located in the lateral subdivision of the central amygdala (CEl), which express protein kinase C-δ (PKC-δ). Channelrhodopsin-2-assisted circuit mapping in amygdala slices and cell-specific viral tracing indicate that PKC-δ$^+$ neurons inhibit output neurons in the medial central amygdala (CEm), and also make reciprocal inhibitory synapses with PKC-δ$^-$ neurons in CEl. Electrical silencing of PKC-δ$^+$ neurons *in vivo* suggests that they correspond to physiologically identified units that are inhibited by the conditioned stimulus, called CEl$_{off}$ units. This correspondence, together with behavioural data, defines an inhibitory microcircuit in CEl that gates CEm output to control the level of conditioned freezing.

The amygdala is a medial temporal lobe region that is central to the acquisition and expression of Pavlovian conditioned fear[1–3]. The amygdala contains multiple anatomically defined nuclei[4]. Associative learning occurs largely in the lateral nucleus, whereas the central nucleus is thought to control the expression of fear[5,6]. However, the central nucleus contains at least three subnuclei (CEm, CEl and capsular central amygdala)[3,4] and multiple neuronal subtypes[7–10] whose role is poorly understood. A full understanding of amygdala function requires its dissection at the cellular level.

Here we have used genetic methods to investigate the functional connectivity and behavioural relevance of a GABAergic subpopulation within CEl[11–13], marked by the expression of PKC-δ. These neurons inhibit output neurons in CEm and also make reciprocal inhibitory connections with PKC-δ$^-$ neurons within CEl. Genetic silencing[14,15] in conjunction with *in vivo* electrophysiology indicates that these neurons probably correspond to 'CEl$_{off}$' units identified in the accompanying paper[16] that are inhibited by the conditioned stimulus (CS). Taken together, these data define the functional connectivity and behavioural relevance of an inhibitory microcircuit within CEl[9,17] that gates output from CEm[12].

Properties of PKC-δ$^+$ neurons in CEl

We sought stable markers for subpopulations in the central nucleus[18,19], which could be used to genetically manipulate their activity *in vivo*[20–22]. PKC-δ marks about 50% of CEl GABAergic neurons (Fig. 1a–d and Supplementary Table 1). These neurons are distinct from those expressing corticotropin-releasing hormone[9] (Fig. 1e–g) or dynorphin[10] (Supplementary Fig. 1a–c), whereas about 40% express enkephalin[9] (Fig. 1h–j and Supplementary Table 1) and about 65% express the oxytocin receptor (Supplementary Fig. 1d–f and Supplementary Table 1), which is implicated in inhibitory gating of CEm[12].

We examined the electrophysiological properties of PKC-δ$^+$ neurons in acute amygdala slices. Because of their low spontaneous activity,

spiking was evoked by the injection of depolarizing current. This analysis confirmed three types of neuron: late-firing, regular spiking and low-threshold bursting neurons[23–25] (Fig. 1k, l and Supplementary Table 2). Fills with neurobiotin and staining with antibodies, as well as recording from fluorescently labelled PKC-δ$^+$ neurons in transgenic mice (see below), indicated that most PKC-δ$^+$ neurons are late-firing (Fig. 1m, o and Supplementary Table 3), whereas the PKC-δ$^-$ population contains both regular spiking and late-firing units (Fig. 1n, o and Supplementary Table 3). Thus, PKC-δ$^+$ neurons have relatively homogeneous electrophysiological properties (Supplementary Table 4; $P < 0.0001$, Fisher's exact test).

Functional connectivity of PKC-δ$^+$ neurons

To gain genetic access to PKC-δ$^+$ neurons, transgenic mice were generated[26] harbouring a bacterial artificial chromosome (BAC) expressing Cre recombinase and the α subunit of a cyan fluorescent protein (CFP)-tagged version of the *Caenorhabditis elegans* glutamate-sensitive chloride channel GluCl (ref. 27) (Fig. 2a), separated by an internal ribosome entry site (ires)[28]. Double labelling for GluClα–CFP and PKC-δ in these PKC-δ:GluClα-ires-Cre mice, as well as crossing to Cre-dependent *lacZ*-expressing reporter mice, revealed neuron-specific expression of the transgene (Fig. 2j–m) that correctly recapitulated the pattern of endogenous PKC-δ expression (Fig. 2b–i).

CEl is known to contain GABAergic neurons that project to CEm[7,12,29,30]. We traced the anterograde projections of PKC-δ$^+$ neurons to CEm by injecting the CEl of the transgenic mice with a Cre-dependent adeno-associated virus (AAV) encoding humanized *Renilla* green fluorescent protein[31] (hrGFP) (Fig. 3a–j). Recombination of injected AAVs was restricted to PKC-δ$^+$ neurons (Supplementary Fig. 2f–j). hrGFP$^+$ fibres derived from CEl projected to CEm (Fig. 3h, inset, and Fig. 3i, j), a result confirmed by retrograde tracing from CEm with the use of the B subunit of cholera toxin (CTB) (Fig. 3k, l). In all, $60.9 \pm 5.7\%$ (mean ± s.e.m., $n = 6$) of CTB$^+$ neurons in CEl were

[1]Division of Biology 216-76, California Institute of Technology, Pasadena, California 91125, USA. [2]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California 91125, USA. [3]Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland. [4]Systems Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [5]Department of Psychology and the Brain Research Institute, University of California, Los Angeles, Los Angeles, California 90095, USA. [6]Laboratory for Neuroimaging, University of California, Los Angeles, Los Angeles, California 90095, USA. [7]Department of Bioengineering, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.
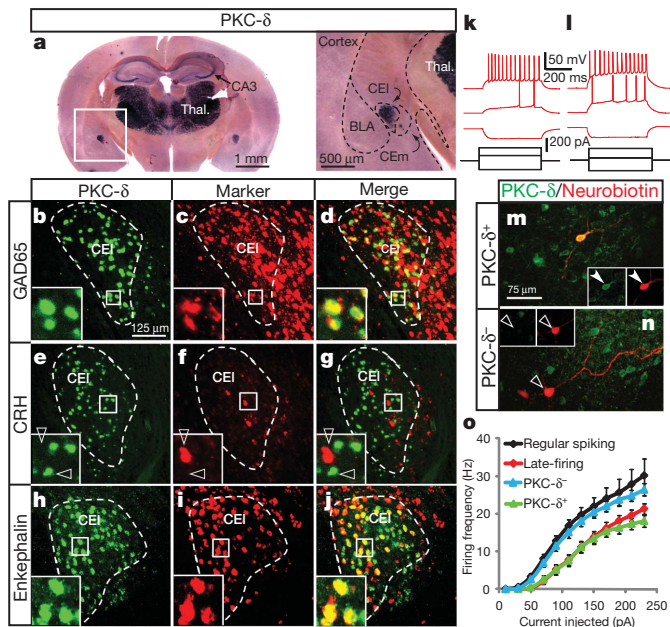
Figure 1 | Characterization of CEl PKC-δ⁺ neurons. a, *In situ* hybridization for PKC-δ mRNA. CA3, hippocampus; thal., thalamus. The boxed area is shown at a higher magnification on the right. **b–j**, Double-label fluorescence *in situ* hybridization for PKC-δ mRNA and the indicated markers. The insets show enlargements of the boxed areas. Filled and open arrowheads indicate doubly labelled and singly labelled cells, respectively. CRH, corticotropin-releasing hormone. **k, l**, Whole-cell patch clamp recordings of late-firing (**k**) and regular spiking (**l**) neurons in acute amygdala slices. Voltage changes (red) produced by the respective current injections (black) are illustrated. Resting membrane potentials were adjusted to about −65 mV. **m, n**, Neurobiotin-filled CEl neurons recorded in **k** and **l**, respectively, after labelling with fluorescent streptavidin and immunostaining for PKC-δ. **o**, Stimulus–response (I/O) curves for neurons based on electrophysiological profile (late-firing, *n* = 22; regular spiking, *n* = 14) or PKC-δ expression (PKC-δ⁺, *n* = 14; PKC-δ⁻, *n* = 12) in CEl. See also Supplementary Tables 2–4. Values are means ± s.e.m.

PKC-δ⁺ (Fig. 3m–o, white arrowhead in inset); the remainder were PKC-δ⁻.

Freezing is controlled by projections from CEm to the peri-aqueductal grey (PAG)[32–34], but CEm contains multiple neuronal subpopulations with different connectivities[34]. To determine whether PKC-δ⁺ neurons make inhibitory synapses onto PAG-projecting CEm output neurons, we combined whole-cell patch clamp recording of retrogradely labelled CEm output neurons, with Cre-dependent optogenetic activation[35] of PKC-δ⁺ neurons (Fig. 3p, q), in acute amygdala slices. Selective expression of channelrhodopsin-2 (ChR2)[36] in PKC-δ⁺ neurons was achieved by injecting a Cre-dependent AAV encoding yellow fluorescent protein (YFP)-tagged ChR2 (ref. 37) into the CEl of PKC-δ:GluClα-ires-Cre transgenic mice (Supplementary Fig. 9a). Whole-cell patch-clamp recordings from ChR2-expressing PKC-δ⁺ neurons (identified using native fluorescence for YFP (ChR2) and CFP (GluClα; Supplementary Fig. 9b)) confirmed that spiking could be induced with about 90% efficiency with 473-nm light flashes at 15 Hz (Supplementary Fig. 9g). Retrograde labelling of CEm projection neurons in the same animals was achieved by injection of Alexa-555-conjugated CTB into the PAG (Fig. 3p), permitting their prospective identification by native fluorescence (Fig. 3q, arrowhead).

In slices prepared from doubly injected animals, optogenetic activation of CEl PKC-δ⁺ neurons elicited robust, picrotoxin-sensitive inhibitory postsynaptic currents (IPSCs) in CEm output neurons (success rate 100%) (Fig. 3r, s; IPSC amplitude 18.3 ± 2.3 pA, *n* = 6 cells) with an average latency of 3.7 ± 0.2 ms (range 2–5 ms, *n* = 30) and a mean temporal jitter of 0.83 ± 0.16 ms (*n* = 6), consistent with mono-synaptic transmission[38]. It also suppressed current-injection-evoked
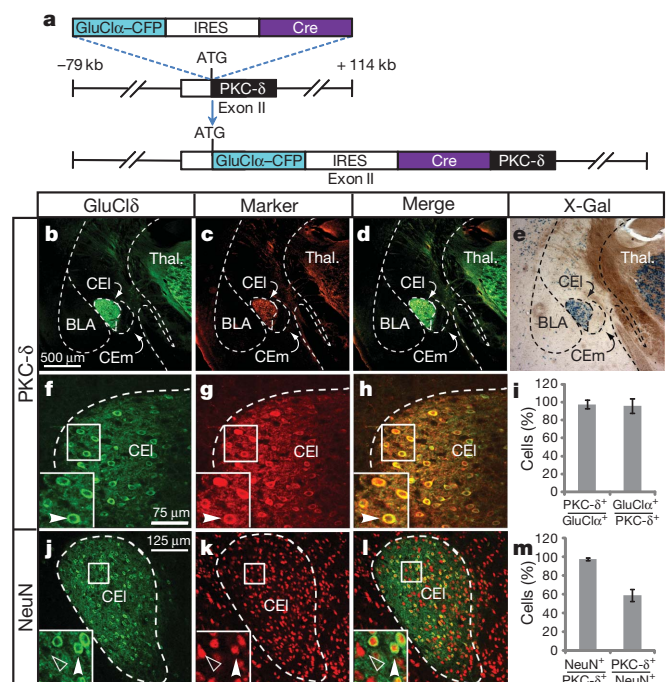
Figure 2 | Transgenic targeting of PKC-δ⁺ neurons. a, Design of PKC-δ::GluClα−CFP-iresCre BAC transgene. kb, kilobases. **b–d, f–h, j–l**, Expression of transgene visualized by double-label immunofluorescent labelling for GFP and the indicated markers. **i, m**, Quantification of **f–h** and **j–l**, respectively. Values are means ± s.e.m. (*n* = 3). **e**, Staining of PKC-δ::GluClα-iresCre; Rosa::loxP-STOP-loxP-lacZ mice with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) reveals β-galactosidase expression (blue) in correct locations. Scale bar in **b** applies to **c–h** and **j–l**.

action potentials in CEm output neurons (Fig. 3t, v) in a picrotoxin-sensitive manner (Fig. 3u, w). Thus, CEl PKC-δ⁺ neurons make inhibitory connections onto PAG-projecting CEm output neurons.

Activation of ChR2-expressing PKC-δ⁺ neurons by light also evoked monosynaptic IPSCs and suppressed current-injection-evoked action potential firing in CEl PKC-δ⁻ neurons (identified by the absence of CFP fluorescence), in a picrotoxin-sensitive manner (Fig. 4a–g). PKC-δ⁺ neurons therefore also make local inhibitory connections within CEl onto PKC-δ⁻ neurons.

To determine whether, conversely, PKC-δ⁺ neurons receive inhibitory input from PKC-δ⁻ neurons, we employed a cell-specific modification of a virally based monosynaptic retrograde tracing method[39]. PKC-δ::GluClα-ires-Cre transgenic mice were first injected with a Cre-dependent AAV encoding both the avian sarcoma/leukosis virus subtype A receptor TVA, and a complementing RV-G protein deleted from the rabies virus strain RV^ΔG (Fig. 4h, left)[40]. Three weeks later, animals were injected in the same site with Env(A)-pseudotyped RV^ΔG virus expressing the red fluorescent protein mCherry (Fig. 4h, right; see Methods). Because Env(A) directs infection exclusively to TVA-expressing cells[39], this manipulation yields expression of mCherry in CEl PKC-δ⁺ neurons in transgenic mice (Fig. 4m–p; inset, open arrowheads) but not in wild-type mice (Fig. 4i, j). Expression of rabies B19 glycoprotein by AAV-infected PKC-δ⁺ neurons permits the trans-synaptic spread of the RV^ΔG virus to input neurons, but no further spread occurs[39]. Monosynaptic retrograde labelling of PKC-δ⁻ neurons occurred extensively in CEl (Fig. 4m, n, p; inset, filled arrowheads), and most labelled neurons were GABAergic (Fig. 4q–t; inset, arrowhead). These data suggest that PKC-δ⁺ neurons receive inhibitory input from PKC-δ⁻ neurons.

PKC-δ⁺ neurons correspond to CEl_off units

Single-unit recording experiments in freely behaving mice have identified two types of CEl units with opposite responses to the CS[16]: units
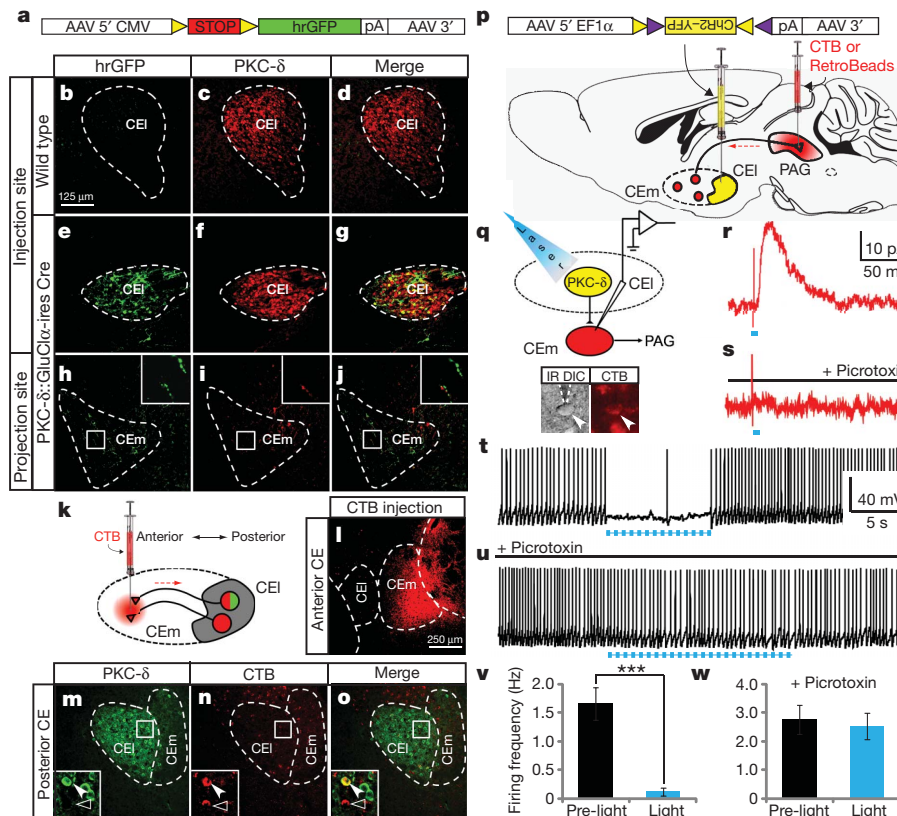
Figure 3 | CEl PKC-δ$^+$ neurons directly inhibit CEm output neurons.
a–j, Anterograde axonal tracing with the use of Cre-dependent hrGFP AAV[31].
k–o, Retrograde tracing from CEm using red fluorescent CTB.
p–w, Optogenetic activation of CEl PKC-δ$^+$ neurons inhibits PAG-projecting neurons in CEm. **p,** Diagram illustrating double-injection/slice-recording experiments. **q,** CTB$^+$ CEm cell (CTB, arrowhead) with patch electrode (infrared differential interference contrast (IR DIC), dashed lines) is illustrated.

r–v, Whole-cell voltage-clamp ($V_{hold} = -40$ mV) (**r, s**) and current-clamp (**t–v**) recordings from a back-labelled CEm neuron. **r, s,** IPSC triggered by a 2-ms 473-nm laser pulse, with (**s**) or without (**r**) 100 μM picrotoxin. **t–v,** Suppression of depolarizing current-injection-evoked spiking by 473-nm laser pulses (2 ms, 15 Hz), with (**u**) or without (**t**) 100 μM picrotoxin. **v, w,** Quantification of data in **t** ($n = 5$ cells; asterisk, $P < 0.001$, t-test) and **v** ($n = 3$ cells; $P = 0.75$, t-test), respectively. Values are means ± s.e.m.

activated by the tone ('CEl$_{on}$' cells), and units inhibited by it ('CEl$_{off}$' cells) (Fig. 5i–l). We wished to determine whether PKC-δ$^+$ neurons correspond to either of these classes of unit. Because the *in vivo* recordings employed extracellular electrodes[16], the cells could not be filled for antibody staining. We therefore examined the effect of reversibly silencing PKC-δ$^+$ neurons on the activity of CEl$_{on}$ and CEl$_{off}$ units, using a mammalian codon-optimized form[41] of the ivermectin-sensitive chloride channel GluCl from *C. elegans*[14,15], mutated to eliminate glutamate sensitivity[27]. Functional expression of this ionotropic receptor requires two subunits, namely GluClα and GluClβ. To restrict the expression of GluClαβ to CEl PKC-δ$^+$ neurons, we employed an intersectional strategy in which GluClα–CFP was expressed transgenically in all PKC-δ$^+$ neurons (Fig. 2b), whereas GluClβ–YFP[25] was expressed in CEl from an AAV vector by stereotaxic injection (Fig. 5a and Supplementary Fig. 2a–e).

We first confirmed silencing of PKC-δ$^+$ neurons by using ivermectin/GluCl in acute amygdala slices from the mice with intersectional GluClαβ expression. Neurons expressing GluClα and/or GluClβ could be prospectively identified by native CFP and YFP fluorescence, respectively (Fig. 5c–e). Bath application of ivermectin decreased the input resistance of cells expressing GluClαβ, but not that of cells expressing either GluClα or GluClβ, from 337 ± 41 to 277 ± 36 MΩ ($n = 4$, $P < 0.04$, paired t-test). Ivermectin also significantly suppressed current-injection-evoked spiking in neurons expressing both subunits (Fig. 5c–h), as well as subthreshold excitatory postsynaptic potentials evoked by stimulation of the lateral nucleus with a bipolar electrode (Supplementary Fig. 3; this bipolar stimulation did not evoke spiking of PKC-δ$^+$ neurons in our slice preparation).

We next examined the effect of silencing PKC-δ$^+$ neurons on CEl$_{on}$ and CEl$_{off}$ unit activity[16], in PKC-δ::GluClα-ires-Cre transgenic mice previously injected in CEl with the AAV::GluClβ virus. CEl$_{on}$ and CEl$_{off}$ units were first identified by CS presentation in conditioned animals, before administration of ivermectin. The spontaneous spiking activity of these units was subsequently measured before, and 3–5 days after, administration of ivermectin. The tonic activity of CEl$_{off}$ units was strongly suppressed after treatment with ivermectin, whereas that of CEl$_{on}$ units was unaffected (Fig. 5m, n). This effect of ivermectin was reversed within 48 h after treatment (Fig. 5m), and was not observed in ivermectin-treated uninjected transgenic animals or in virally injected animals not treated with ivermectin (Fig. 5n). Silencing of PKC-δ$^+$ neurons with ivermectin/GluCl also increased CEm unit activity (Fig. 5m, n, red symbols), which is consistent with our observation that PKC-δ$^+$ CEl neurons inhibit CEm output neurons (Fig. 3q–w). The simplest interpretation of these data is that CEl$_{off}$ units are PKC-δ$^+$ neurons, although this does not necessarily imply the converse.

Behavioural effect of silencing CEl PKC-δ$^+$ neurons

Finally, we tested the effect of suppressing PKC-δ$^+$ neuronal activity on fear conditioning (see Methods). We initially investigated the effect of silencing during both training and testing, analogous to a pre-training lesion. Although ivermectin at the dose used (10 mg kg^{-1}) did not significantly affect freezing in wild-type animals (Supplementary Fig. 4a), to avoid potential interactions between treatment with ivermectin and viral infection, in most experiments we also treated single subunit-expressing control animals (transgenic or wild-type animals injected
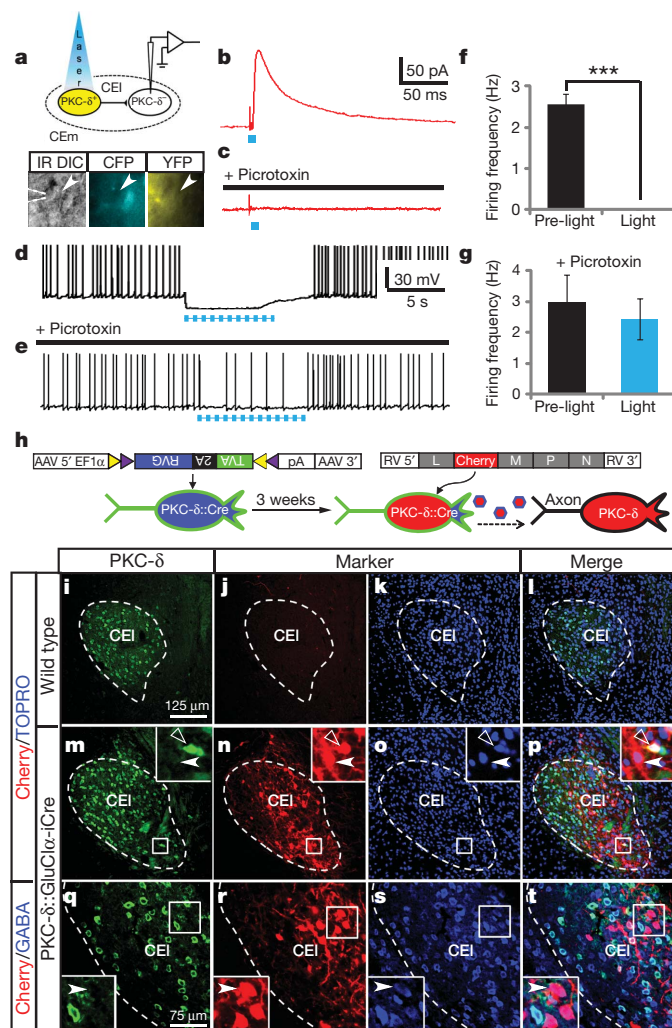
Figure 4 | PKC-δ⁺ and PKC-δ⁻ make reciprocal inhibitory connections in CEl. **a**, Diagram and micrograph from slice preparation. Arrowhead indicates PKC-δ (CFP)⁻, ChR2 (YFP)⁻ recorded cell in CEl. **b–e**, Whole-cell voltage-clamp ($V_{hold} = -40$ mV) (**b**, **c**) and current-clamp (**d**, **e**) recordings from a PKC-δ⁻ CEl neuron showing light-triggered IPSCs (**b**), picrotoxin-sensitive IPSCs (**c**) or picrotoxin-sensitive (**e**) suppression of depolarizing current-injection-evoked spiking (**d**) by ChR2 activation (blue dots; 2 ms, 15 Hz). **f**, Quantification of data in **d** ($n = 5$ cells; three asterisks, $P < 0.001$, t-test). **g**, Quantification of data in **e** ($n = 3$ cells; $P = 0.66$, t-test). **h**, Diagram illustrating cell-specific rabies virus infection. **i–p**, Immunostaining for PKC-δ (**i**, **m**), intrinsic mCherry fluorescence (**j**, **n**) and nuclear staining with TOPRO-3 (**k**, **o**) 3 days after injection of transgenic (**m–p**) and wild-type (**i–l**) mice with RV^ΔG. Primary infected PKC-δ⁺ neurons (insets in **m** and **n**, open arrowheads) and retrogradely labelled PKC-δ⁻ cells (insets in **m** and **n**, filled arrowheads) are indicated. **q–t**, Triple labelling for PKC-δ (**q**), GABA (**s**) and mCherry (**r**). mCherry-labelled PKC-δ⁻ cells are GABAergic (inset, arrowhead). Values are means ± s.e.m.

with GluClα or GluClβ virus, respectively) with ivermectin. These control groups were not significantly different from each other (Supplementary Fig. 4b) and were pooled for statistical analysis.

In some experiments (five to eight animals per group), when we used a high-titre (10^{13} ml⁻¹) preparation of AAV2:GluClβ (ref. 14), a significantly higher level of freezing was observed in experimental animals than in controls; however, in other experiments with different virus preparations, no difference between groups was detected (Supplementary Note 1). Quantitative histological analysis (Supplementary Fig. 10) revealed considerable variation in the level and bilaterality of GluClβ–YFP expression among individual animals (Supplementary Fig. 5a–c). We therefore pooled data from multiple experiments (total $n = 54$ animals/group) for statistical analysis. A randomized-block analysis of variance (ANOVA)[42] with six experimental and six control subjects assigned to each of nine blocks based on infection rate (total $n = 108$; Supplementary Table 5 and Supplementary Fig. 5d–f) indicated a significant block × group interaction during both the CS presentation ($F_{(8,90)} = 2.298$, $P < 0.05$ by post-hoc Bonferroni t-test) and post-CS periods ($F_{(8,90)} = 2.459$, $P < 0.05$) but not during the baseline period ($F_{(8,90)} = 1.41$, $P = 0.205$). Freezing was significantly higher in the experimental group only in the block with the highest infection rate (Table 1, Supplementary Table 5 and Supplementary Fig. 5e, f). Freezing in the experimental group was also significantly higher than in controls (including GluClαβ animals tested without ivermectin; $n = 6$) among animals expressing GluClβ bilaterally in the central nucleus at a level above the median infection rate (Supplementary Fig. 7). There was no significant decrease in baseline locomotor activity or in the activity burst produced by the first presentation of the unconditioned stimulus during training (Supplementary Fig. 8) in experimental animals, indicating that the increased freezing was not a consequence of either decreased locomotor activity or increased sensitivity to the unconditioned stimulus[13]. Freezing levels during the pre-training and pre-test baseline were unaffected by ivermectin in experimental animals (Supplementary Table 5, Supplementary Figs 5 and 7, and data not shown).

Discussion

We have used genetically based methods to investigate the functional connectivity of a subpopulation of CEl GABAergic neurons identified by the expression of PKC-δ. Our data suggest that these neurons participate in a recurrent inhibitory circuit within CEl that inhibits CEm output to brainstem centres that control freezing, consistent with earlier suggestions[9,12]. In the accompanying paper[16] two populations of CEl units are identified in vivo with opposite responses to the CS. The data presented here strongly suggest that CEl_off units are PKC-δ⁺ neurons, linking molecular and physiological identity. Because PKC-δ⁺ neurons pre-exist in untrained animals, this linkage suggests that CEl_off and CEl_on units, which are robustly identifiable after conditioning (Fig. 5k, l), arise through plasticity-dependent changes in deterministic CEl subpopulations rather than by stochastic selection from a population of initially homogeneous cells.

The link between PKC-δ⁺ neurons and CEl_off units is also consistent with complementary connectivity data in the two studies. For example, it has been shown that in vivo the activity of CEl_off units is inversely correlated with the activity of CEl_on units, suggesting reciprocal inhibition[16]. Our channelrhodopsin-assisted circuit mapping[38] and cell-specific

Table 1 | Behavioural effect of silencing PKC-δ⁺ neurons in CEl

Group (block 9)	Infection rate (%)	Baseline freezing (%)	CS freezing (%)	Post-CS freezing (%)
Control (GluClα or GluClβ + ivermectin)	40.0 ± 0.8	5.8 ± 2.2	49.4 ± 10.4	19.2 ± 6.0
Experimental (GluClαβ + ivermectin)	40.6 ± 3.5	2.6 ± 1.5 (n.s.)	91.1 ± 4.7 ($P < 0.05$)	75.3 ± 8.4 ($P < 0.001$)

The data are derived from the randomized-block ANOVA in Supplementary Table 5 (total $n = 108$; six experimental and six control animals assigned to each of nine blocks), and illustrate the block with the highest level of viral infection in CEl (infection rate: the percentage of PKC-δ⁺ cells expressing GluClβ–YFP; see Supplementary Fig. 10). The control group was pooled (Supplementary Fig. 4b) single-subunit-expressing mice (GluClα or GluClβ alone) treated with ivermectin. Values are means ± s.e.m.; n.s., not significantly different. P values are from a post-hoc Bonferroni t-test. A significant linear component to the block × group interaction indicated that the difference between groups tended to increase with infection rate ($F_{(1,90)} = 22.98$, $P < 0.0001$ for CS; $F_{(1,90)} = 8.85$, $P < 0.006$ for post-CS). See also Supplementary Table 5 and Supplementary Fig. 5. Control GluClαβ-expressing mice not treated with ivermectin were significantly different from the experimental group among animals showing bilateral AAV infection above the median rate ($n = 6–21$; Supplementary Fig. 7).
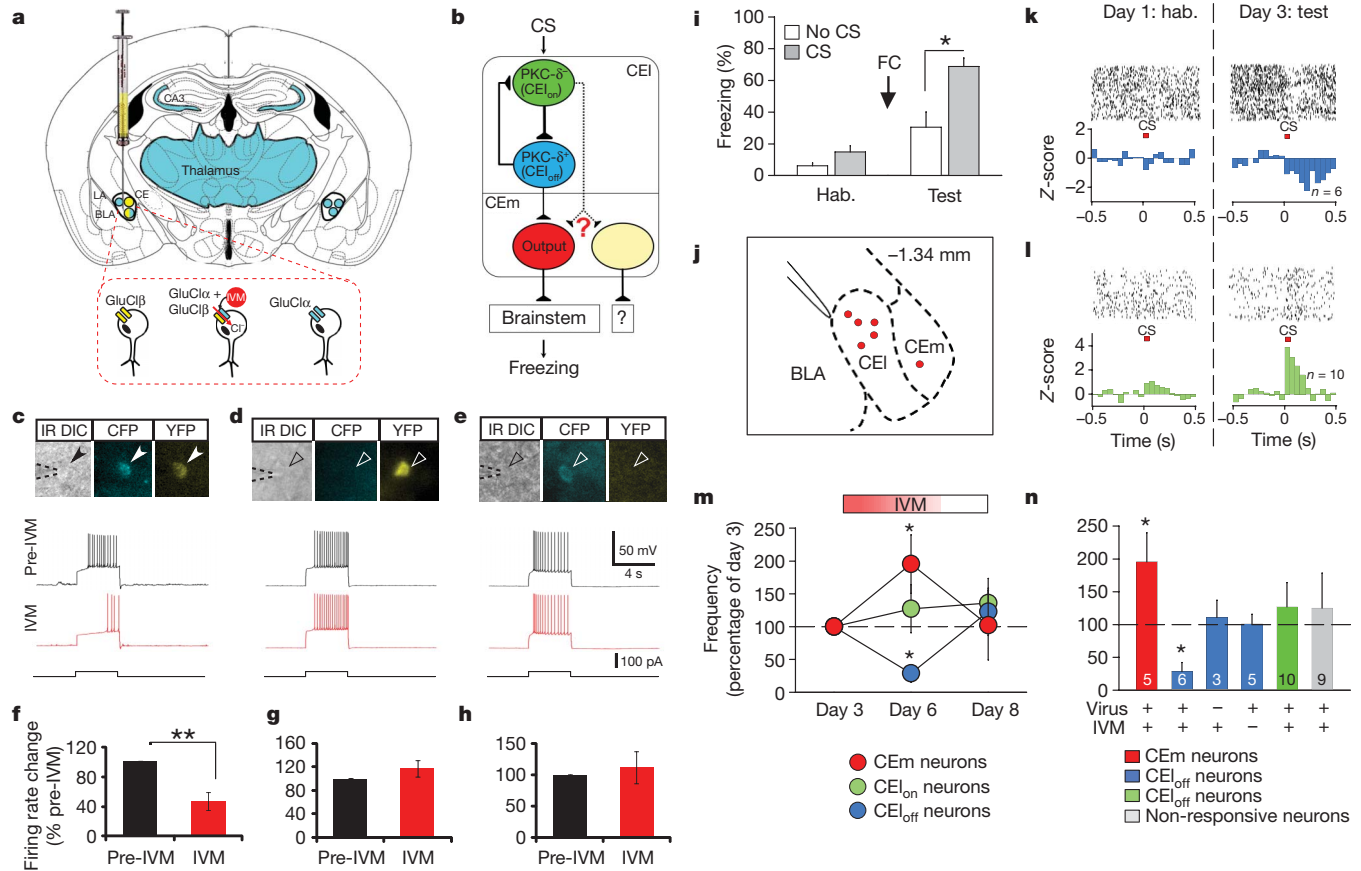
Figure 5 | CEl_off neurons are PKC-δ⁺. **a**, Selective expression of GluClαβ in CEl PKC-δ⁺ neurons. Yellow, AAV::GluClβ–YFP; blue, PKC-δ::GluClα–CFP transgene. **b**, Inhibitory connections in CE[16]. **c–e**, PKC-δ⁺ cells expressing GluClαβ (**c**), GluClβ–YFP (**d**) or GluClα–CFP (**e**). Upper and middle traces, spiking before and after ivermectin (IVM) (20 nM), respectively; bottom trace, current injection. **f–h**, Quantification of **c–e** (**f**, $n = 4$ cells, $P = 0.005$; **g**, $n = 5$ cells, $P = 0.73$; **h**, $n = 4$ cells, $P = 0.66$; paired t-tests). **i–n**, Silencing PKC-δ⁺ neurons during chronic *in vivo* recording. **i**, Behavioural data for recorded mice. Hab., habituation; FC, fear conditioning ($n = 5$). **j**, Coronal diagram showing recording sites. BLA, basolateral amygdala. **k**, **l**, Representative raster plots (upper) and normalized population peristimulus time histograms (lower) of CEl_off (**k**; lower, $n = 6$ neurons, three mice) and CEl_on neurons (**l**; lower, $n = 10$ neurons, four mice) showing CS-evoked changes ($P < 0.05$). For unit isolation see Supplementary Fig. 11. **m**, Tonic activity of CEl_off (one-way ANOVA ($F_{(2,15)} = 4.845$, $P = 0.024$) with post-hoc Bonferroni t-test (asterisk, $P < 0.05$)), but not of CEl_on neurons (one-way ANOVA ($F_{(2,27)} = 0.391$, $P = 0.680$)), is significantly (day 6) and reversibly (day 8) decreased by ivermectin, whereas CEm activity is increased ($n = 5$ units, one mouse; one-way Kruskal–Wallis ANOVA ($H = 7.487$, $P = 0.024$) with post-hoc Tukey's Honestly Significant Difference (asterisk, $P < 0.05$)). **n**, Specificity controls for **m**. Asterisk, $P < 0.05$ (paired t-tests, treatment versus pre-treatment values)). Ivermectin injection without virus ($n = 3$; $P = 0.765$); AAV2::GluClβ virus without ivermectin ($n = 5$; $P = 0.940$). CEl_on neurons ($n = 10$; $P = 0.497$); CS-non-responsive neurons ($n = 9$; $P = 0.644$). Error bars show s.e.m.

monosynaptic retrograde viral tracing[39] data directly demonstrate mutual inhibitory connections between PKC-δ⁺ and PKC-δ⁻ neurons in CEl. It has also been shown that exposure to CS both inhibits CEl_off unit spiking and increases CEm unit activity[16], with a short latency implying an inhibitory connection. Our data directly demonstrate inhibitory synapses from CEl PKC-δ⁺ neurons onto brainstem-projecting CEm output neurons, and indicate that silencing the former increases the tonic activity of the latter *in vivo*. Thus, our direct mapping of PKC-δ⁺ synaptic connectivity in amygdala slices is consistent with inferences of CEl_off unit connectivity based on correlative *in vivo* recordings (Fig. 5b).

Consistent with these electrophysiological and connectional data, genetic silencing of PKC-δ⁺ neurons yielded a statistically significant enhancement of conditional freezing (during both the CS-on and post-CS periods) among those animals with the highest level of AAV::GluClβ infection in CEl. Nevertheless, these behavioural results should be interpreted with caution, given the variability in results between individual experiments (see Supplementary Note 1). Furthermore, although treatment with ivermectin caused a decrease in CEl_off tonic spiking activity in GluClαβ-expressing animals, how this effect influences CS-evoked freezing, which is correlated with phasic CS response in CEl_off unit activity[16], is not yet clear.

Although the CEl circuitry defined here and in ref. 16 is remarkably consistent, the behavioural results in the two studies differ in some respects. For example, muscimol injection into CEl evoked freezing in unconditioned animals[16], whereas selective silencing of PKC-δ⁺ neurons with ivermectin/GluCl did not. Furthermore, transient inhibition of CEl during training with muscimol attenuated fear conditioning[16], whereas chronic silencing of PKC-δ⁺ neurons during both training and testing did not. We have not yet examined the behavioural effects of silencing PKC-δ⁺ neurons during training compared with testing, and this will be a significant question for future studies. Differences between the consequences of the inhibition of CEl by muscimol and those arising from the silencing of CEl PKC-δ⁺ neurons by ivermectin/GluCl may reflect differences in the cellular specificity, efficiency or time course of the two types of manipulation (Supplementary Note 2).

The CEl/CEm microcircuit revealed by these companion studies has properties worthy of further investigation. For example, the mutual inhibitory connections between CEl_off and CEl_on units could provide a positive feedback loop to amplify CS-evoked activity in CEl_on units (Fig. 5b)[43]. These antagonistic connections could also potentially encode different states, through stable imbalances in tonic activity[16] or different oscillatory regimes, that could influence the learning and/or expression of conditioned fear. A currently unresolved paradox is

why, if CEl$_{on}$ units (like CEl$_{off}$ units) send inhibitory projections to CEm[16], the activation of CEl$_{on}$ units by the CS results in increased rather than decreased activity of CEm output neurons (see Supplementary Note 3). Resolution of this paradox will require determining whether these two CEl populations target the same or different classes of CEm neuron[34], and the relative strength of these connections. Genetically based synaptic tracing and functional manipulations of CEl$_{on}$ units should clarify this issue, and should also provide direct tests of their causal role in conditioned freezing and other emotional behaviours. The ability to identify and selectively manipulate PKC-δ^+ and other molecularly defined CEl subpopulations prospectively[44,45] should also open the way to investigating their roles in animal models of phobic or anxiety disorders, and in the mechanism of action of drugs used to treat such disorders[46].

METHODS SUMMARY

Histochemical methods. Single-colour and double-label fluorescence *in situ* hybridizations were performed on fresh frozen sections, using RNA probes labelled with digoxigenin (Roche) and with 2,4-dinitrophenol (PerkinElmer). Immunofluorescence was performed on cryosections of tissue perfused with 4% paraformaldehyde, following standard protocols.

Generation of PKC-δ::GluClα-ires-Cre transgenic mice. A GluClα-ires-Cre cassette was inserted into PKC-δ BAC clone RP23-283B12 (CHORI). The modified BAC was injected into FVB mouse embryos (GENSAT). Transgenic founders were backcrossed ($n > 5$) to C57Bl6/N. These mice are available through GENSAT[26].

Pharmacogenetic silencing *in vitro* and *in vivo*. PKC-δ::GluClα-ires-Cre transgenic or wild-type mice were injected with about 10^9 particles of AAV::GluClβ (or, in control experiments, AAV::GluClα) into the central amygdala and allowed to recover for 4 weeks. For fear conditioning experiments, mice were habituated and then injected with ivermectin ($10\,mg\,kg^{-1}$; Phoenix) on day 1. On day 2, animals received tone/foot-shock pairings in context A (Coulbourn, Med Associates). On day 3, (and for *in vivo* recordings on days 6 and 8), animals were placed in context B and freezing was scored before (baseline), during and after CS presentations. For *in vitro* recordings, acute brain slices were prepared and superfused with 20 nM ivermectin in artificial cerebrospinal fluid to induce pharmacogenetic silencing. Single-unit recordings were performed in freely behaving animals as described[16].

Optogenetic circuit dissection. PKC-δ::GluClα-ires-Cre transgenic mice were injected with 10^9 particles of Cre-dependent ChR2 AAV into the central amygdala, and (in some animals) CTB into the PAG for retrograde labelling. Four weeks later, neurons were stimulated with light, in acute brain slices, through a 200-μm optical fibre (Thorlabs) emitting 5–10 mW of 473-nm laser light (Crystalaser).

Virus based trans-synaptic tracing. PKC-δ::GluClα-ires-Cre transgenic mice were injected sequentially into the central amygdala with 10^4 particles of AAV encoding Cre-dependent TVA and rabies B19 glycoprotein. Three weeks later, animals were injected in the same site with 10^5 particles of RV$^{\Delta G}$ rabies virus pseudotyped with EnvA, and analysed one week later. All animal experiments were conducted under protocols approved by the Caltech Institutional Animal Care and Use Committee (IACUC) and the Salk Institute Biosafety Committee.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Davis, M., Walker, D. L. & Myers, K. M. Role of the amygdala in fear extinction measured with potentiated startle. *Ann. NY Acad. Sci.* **985**, 218–232 (2003).
2. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
3. Paré, D., Quirk, G. J. & LeDoux, J. E. New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* **92**, 1–9 (2004).
4. Pitkänen, A., Savander, V. & LeDoux, J. E. Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends Neurosci.* **20**, 517–523 (1997).
5. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
6. Medina, J. F., Repa, J. C., Mauk, M. D. & LeDoux, J. E. Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Rev. Neurosci.* **3**, 122–131 (2002).
7. Cassell, M. D., Freedman, L. J. & Shi, C. The intrinsic organization of the central extended amygdala. *Ann. NY Acad. Sci.* **877**, 217–241 (1999).
8. Cassell, M. D., Gray, T. S. & Kiss, J. Z. Neuronal architecture in the rat central nucleus of the amygdala: a cytological, hodological, and immunocytochemical study. *J. Comp. Neurol.* **246**, 478–499 (1986).
9. Day, H. E. W., Curran, E. J., Watson, S. J. & Akil, H. Distinct neurochemical populations in the rat central nucleus of the amygdala and bed nucleus of the stria terminalis: evidence for their selective activation by interleukin-1β. *J. Comp. Neurol.* **413**, 113–128 (1999).
10. Marchant, N. J., Densmore, V. S. & Osborne, P. B. Coexpression of prodynorphin and corticotrophin-releasing hormone in the rat central amygdala: evidence of two distinct endogenous opioid systems in the lateral division. *J. Comp. Neurol.* **504**, 702–715 (2007).
11. Ehrlich, I. *et al.* Amygdala inhibitory circuits and the control of fear memory. *Neuron* **62**, 757–771 (2009).
12. Huber, D., Veinante, P. & Stoop, R. Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. *Science* **308**, 245–248 (2005).
13. Wilensky, A. E., Schafe, G. E., Kristensen, M. P. & LeDoux, J. E. Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J. Neurosci.* **26**, 12387–12396 (2006).
14. Lerchner, W. *et al.* Reversible silencing of neuronal excitability in behaving mice by a genetically targeted, ivermectin-gated Cl⁻ channel. *Neuron* **54**, 35–49 (2007).
15. Slimko, E. M., McKinney, S., Anderson, D. J., Davidson, N. & Lester, H. A. Selective electrical silencing of mammalian neurons *in vitro* by the use of invertebrate ligand-gated chloride channels. *J. Neurosci.* **22**, 7373–7379 (2002).
16. Ciocchi, S. *et al.* Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* doi:10.1038/nature09559 (this issue).
17. Day, H. E., Nebel, S., Sasse, S. & Campeau, S. Inhibition of the central extended amygdala by loud noise and restraint stress. *Eur. J. Neurosci.* **21**, 441–454 (2005).
18. Zirlinger, M. & Anderson, D. Molecular dissection of the amygdala and its relevance to autism. *Genes Brain Behav.* **2**, 282–294 (2003).
19. Zirlinger, M., Kreiman, G. & Anderson, D. J. Amygdala-enriched genes identified by microarray technology are restricted to specific amygdaloid subnuclei. *Proc. Natl Acad. Sci. USA* **98**, 5270–5275 (2001).
20. Callaway, E. M. A molecular and genetic arsenal for systems neuroscience. *Trends Neurosci.* **28**, 196–201 (2005).
21. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits. *Neuron* **57**, 634–660 (2008).
22. Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. & Deisseroth, K. Circuit-breakers: optical technologies for probing neural signals and systems. *Nature Rev. Neurosci.* **8**, 577–581 (2007).
23. Chieng, B. C., Christie, M. J. & Osborne, P. B. Characterization of neurons in the rat central nucleus of the amygdala: cellular physiology, morphology, and opioid sensitivity. *J. Comp. Neurol.* **497**, 910–927 (2006).
24. Schiess, M. C., Callahan, P. M. & Zheng, H. Characterization of the electrophysiological and morphological properties of rat central amygdala neurons *in vitro*. *J. Neurosci. Res.* **58**, 663–673 (1999).
25. Lopez de Armentia, M. & Sah, P. Firing properties and connectivity of neurons in the rat lateral central nucleus of the amygdala. *J. Neurophysiol.* **92**, 1285–1294 (2004).
26. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
27. Li, P., Slimko, E. M. & Lester, H. A. Selective elimination of glutamate activation and introduction of fluorescent proteins into a *Caenorhabditis elegans* chloride channel. *FEBS Lett.* **528**, 77–82 (2002).
28. Wagstaff, M. J. *et al.* Gene transfer using a disabled herpes virus vector containing the EMCV IRES allows multiple gene expression *in vitro* and *in vivo*. *Gene Ther.* **5**, 1566–1570 (1998).
29. Veinante, P. & Freund-Mercier, M. J. Branching patterns of central amygdaloid nucleus afferents in the rat: single axon reconstructions. *Ann. NY Acad. Sci.* **985**, 552–553 (2003).
30. Sun, N., Yi, H. & Cassell, M. D. Evidence for a GABAergic interface between cortical afferents and brainstem projection neurons in the rat central extended amygdala. *J. Comp. Neurol.* **340**, 43–64 (1994).
31. Gautron, L., Lazarus, M., Scott, M. M., Saper, C. B. & Elmquist, J. K. Identifying the efferent projections of leptin-responsive neurons in the dorsomedial hypothalamus using a novel conditional tracing approach. *J. Comp. Neurol.* **518**, 2090–2108 (2010).
32. De Oca, B. M., De Cola, J. P., Maren, S. & Fanselow, M. S. Distinct regions of the periaqueductal gray are involved in the acquisition and expression of defensive responses. *J. Neurosci.* **18**, 3426–3432 (1998).
33. Kim, J. J., Rison, R. A. & Fanselow, M. S. Effects of amygdala, hippocampus, and periaqueductal grady lesions on short- and long-term contextual fear. *Behav. Neurosci.* **107**, 1093–1098 (1993).
34. LeDoux, J. E., Iwata, J., Cicchetti, P. & Reis, D. J. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.* **8**, 2517–2529 (1988).
35. Kravitz, A. V. *et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).
36. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007).
37. Cardin, J. A. *et al.* Targeted optogenetic stimulation and recording of neurons *in vivo* using cell-type-specific expression of Channelrhodopsin-2. *Nature Protocols* **5**, 247–254 (2010).
38. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).
39. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).

40. Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nature Methods* **4,** 47–49 (2007).
41. Slimko, E. M. & Lester, H. A. Codon optimization of *Caenorhabditis elegans* GluCl ion channel genes for mammalian cells dramatically improves expression levels. *J. Neurosci. Methods* **124,** 75–81 (2003).
42. Edwards, A. L. *Experimental Design in Psychological Research* 4th edn, 249–251 (Holt, Reinhard & Winston, 1972).
43. Wickens, J. R., Arbuthnott, G. W. & Shindou, T. Simulation of GABA function in the basal ganglia: computational models of GABAergic mechanisms in basal ganglia function. *Prog. Brain Res.* **160,** 313–329 (2007).
44. Gozzi, A. *et al.* A neural switch for active and passive fear. *Neuron* **67,** 656–666 (2010).
45. Tsetsenis, T., Ma, X. H., Lo Iacono, L., Beck, S. G. & Gross, C. Suppression of conditioning to ambiguous cues by pharmacogenetic inhibition of the dentate gyrus. *Nature Neurosci.* **10,** 896–902 (2007).
46. Ressler, K. J. & Mayberg, H. S. Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nature Neurosci.* **10,** 1116–1124 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions W.H. initiated the project, generated BAC constructs, designed experiments, performed anatomical, viral injections and behavioural experiments, and wrote the manuscript. P.S.K. contributed to experimental design, performed viral injections, behavioural experiments, data analysis and interpretation. H.C. contributed to experimental design and performed viral injections and slice electrophysiology experiments, data analysis and interpretation. S.C. and A.L. designed, performed and interpreted *in vivo* recording experiments (Fig. 5). N.R.W. and E.M.C. performed rabies virus injection experiments. R.P. performed supplementary behavioural experiments and M.S.F. contributed to their interpretation and to statistical analysis. J.B. and H.-W.D. performed supplementary stereotaxic viral injection experiments. K.D. provided Cre-dependent ChR2 constructs and advice. D.J.A. conceived the project, contributed to experimental design and interpretation and wrote the manuscript. P.S.K., H.C. and S.C. contributed equally. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J.A. (wuwei@caltech.edu).

METHODS

Generation of PKC-δ::GluClα-iCre transgenic mice. The PKC-δ::GluClα-iCre targeting construct was assembled in pGEM-T Easy (Promega) by PCR-cloning nucleotides −425 to −1 (+1 corresponding to the PKC-δ start codon, primer set 5′-ACACACCGCGGCGCGCCCTAAAGAGGCAGGAGGCATGTG-3′ and 5′-CCATGATGGAGCCTGGAGTGAG-3′) and +4 to +561 (primer set 5′-TCTCTGCTAGCCCGGGACCCTTCCTGCGCATCTC-3′ and 5-TGTGTG GTCGACTTAATTAAACTAGTGACCTTTCCAGCCATCACGTG-3′) of PKC-δ genomic sequences 5′ and 3′ to the GluClα open reading frame (ORF)[41], using *Ksp*I-blunt/*Mlu*NI and *Nhe*I–*Sal*I restriction sites, respectively. The resulting PKC-δ::GluClα cassette was then cloned into the pLD53 shuttle vector[47]. An additional expression cassette containing the encephalomyocarditis virus internal ribosome entry site (IRES) followed by an ORF coding for Cre-recombinase was cloned into the *Nhe*I site immediately downstream of GluClα. This vector was then used to introduce the GluClα-iCre expression cassette into a BAC clone containing the complete gene encoding PKC-δ (RP23-283B12) by Rec A-mediated homologous recombination in bacteria[47]. Homologous recombination was verified by PCR (5′ arm primer set 5′-AGACCAGGGTAGGAGTCGGTG-3′ and 5′-GATCAGG GAAGCGATGATCAG-3′; 3′ arm primer set 5′-GAGACCAAGACCGAGT GGAA-3′ and 5′-CACAGGTTAGCCATGACCTG-3′) and Southern blotting (5′ arm probe primer set 5′-TGTTCATGGGGTTTCTCACAG-3′ and 5′-ACCG ACTCCTACCCTGGTCAG-3′; 3′ arm probe primer set 5′-AGGTCATGGC TAACCTGTGG-3′ and 5′-GGCAGAGAAGTCAGACTGGG-3′). Transgenic mice were generated by pronuclear injection of CsCl-prepared BAC DNA linearized with P-*Sce*I in FVB embryos[26], which yielded two independent transgenic lines with indistinguishable expression patterns. One of them was backcrossed for more than five generations to C57Bl6/N. Germline transmission and genotyping of transgenic offspring were traced by PCR on genomic tail DNA (primer set 5′-GCTACA TCAAGGCCATCGAC-3′ and 5′-AACTCCAGCAGGACCATGTGATCG-3′). The line is available through GENSAT.

***In situ* hybridizations.** *In situ* hybridization was performed on fresh frozen brain sections 20 µm thick.

Processing of sections, hybridization and probe detection for single colour *in situ* hybridizations were performed with standard protocols[19], with the following specifications: probes were cloned from whole mouse brain cDNA library (Invitrogen; GAD65 primer set 5′-ATCTCCAACATGTATGCCATGCTCATT GCC-3′ and 5′-TTACAAATCTTGTCCGAGGCGTTCGA-3′; corticotropin-releasing hormone primer set 5′-AACGGAGTAAGGGCAGGAATGGAGACA GAG-3′ and 5′-GTTGCTGTGAGCTTGCTGAGCTAACTGCTCTGC-3′; enkephalin primer set 5′-TAGGGTCCAAGCTCTCATTGAGGCACCCGG-3′ and 5′-GCTTCAGAACCGCATAAAGCCCCCGTAT-3′; oxytocin receptor 5′-CTG GCCCACCAGGCCAGCCGCTGGGTGGTG-3′ and 5′-AATCCCCATCTCCT TGGGGAATTTTAGGAT-3′) transcribed from linearized templates in pTeasy (Promega) or pCR2.1 (Invitrogen) using Sp6 polymerase (9PIP108; Promega) or T7 polymerase (9PIP207; Promega) and digoxigenin-labelled nucleotide mix (11277073910; Roche), used at a final concentration of 1 µg ml^{-1}, and detected with anti-digoxigenin–horseradish peroxidase, Fab fragments (1-207-733; Roche).

Processing of sections and hybridization for double-label fluorescence *in situ* hybridization were performed as above, with the following specifications: digoxigenin and 2,4-dinitrophenol-UTP (NEL555001; PerkinElmer) probe pairs were used at a final concentration of 1 µg ml^{-1} each. For detection, sections were each blocked first in 20% sheep serum in 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween 20 and NEN-blocking solution (PerkinElmer) for 30 min. Digoxigenin-labelled probe was detected with anti-digoxigenin horseradish peroxidase (Roche) (1:500 dilution, 2 h), biotin-tyramide (1:100 dilution, 8 min, room temperature (20–25 °C); PerkinElmer), Vecta Stain Elite ABC Kit (PK-6100; Vector) and Cy3-Tyramide (Cy3 tyramide NEL704A; 1:50 dilution, 1 h; PerkinElmer). Horseradish peroxidase was inactivated by incubation for 1 h in 3% H_2O_2 in PBS, followed by heat denaturation at 85 °C for 5 min in 10 mM tris(hydroxymethyl) aminomethane, 1 mM EDTA pH 7.5. Sections were blocked again as above and 2,4-dinitrophenol (DNP)-labelled nucleotides were detected with anti-DNP horseradish peroxidase (1:500, 2h), DNP-tyramide, anti-DNP horseradish peroxidase (1:500 dilution, 2 h, all TSA DNP Kit, NEL747A; PerkinElmer) and fluorescein-tyramide (1:100 dilution, NEL701; PerkinElmer). Sections were mounted in Fluoro Gel (17985-10; Electron Microscopy Sciences). All washing steps and incubations were performed in accordance with the respective manufacturers' recommendations. Sections were viewed under a Zeiss Axioscope microscope and a Leica TCS SP confocal microscope.

Immunofluorescent labelling. For immunofluorescent labelling, mice were transcardially perfused with 4% paraformaldehyde in PBS. Brains were removed and cryoprotected for 16 h at 4 °C in 15% sucrose. Cryo-sections 20–30 µm thick were dried in air for 30 min and then rehydrated in PBST (PBS plus 0.1% Triton X-100). In some instances, immunoreactivity was increased by incubating the slides in 10 mM sodium citrate, 0.05% Tween 20 pH 6 for 10 min at 95 °C. Non-specific binding was blocked with 1% BSA in PBST for 30 min. Primary antibodies (mouse anti-PKC-δ (610398; BD Biosciences), guinea-pig anti-dynorphin (GP10110; Neuromics), rabbit anti-GFP (A11122; Invitrogen), rabbit anti-hrGFP (240142; Stratagene), mouse anti-NeuN (MAB377; Chemicon), goat anti-CTB (703; List) diluted 1:300–1:1,000 in blocking solution, were incubated for 16 h at 4 °C. Standard secondary antibodies (Invitrogen) in blocking solution were incubated for 3 h at room temperature. Unbound primary and secondary antibodies were each washed by incubation three times in PBST for 10 min. Sections were mounted in Fluoro Gel (17985-10; Electron Microscopy Sciences) and viewed under a Zeiss Axioscope microscope and a Leica TCS SP confocal microscope.

Animal maintenance. Animals were group housed at 23 °C with *ad libitum* access to food and water in a 13-h day/11-h night cycle, with the day starting at 07:00.

Histological analysis. All histological quantifications (Fig. 2i, m, Supplementary Figs 2e, j, 5, 7 and 10, and Supplementary Tables 1 and 5) are based on at least three coronal sections spaced equidistantly along the rostro-caudal axis of the central amygdala.

The intrinsic fluorescence of GluClα−CFP or GluClβ–YFP can be readily detected in live brain slices (Fig. 4a and 5c–e and Supplementary Fig. 9a, b). In the perfusion-fixed tissue used for all histological analysis, however, the fluorescent signal is substantially weaker. GFP immunofluorescent labelling directed against its CFP or YFP tag was therefore used wherever possible for more accurate results (Fig. 1b–h, j–l and Supplementary Figs 2b–d and 5). However, this immunohistochemical method could not be used to quantify the extent of infection by AAV::GluClβ–YFP in PKC-δ::GluClαCFP-iCre transgenic mice, because of cross-reactivity of the anti-GFP antibody with the transgene-encoded CFP. The fraction of virally infected PKC-δ^{+} neurons in PKC-δ::GluClαCFP-iCre transgenic brains was therefore estimated by quantifying the number of cells expressing intrinsic YFP fluorescence (which could be spectrally distinguished from the endogenous CFP fluorescence of the GluClα transgene). However, because this intrinsic fluorescence signal underestimates the extent of viral infection in perfusion-fixed tissue (see above), we constructed a standard curve by using wild-type mice infected with AAV:GluClβ–YFP virus, in which the number of infected cells measured by using native YFP fluorescence was plotted against the number of infected PKC-δ^{+} cells measured by using double-immunofluorescence staining with anti-GFP and anti-PKC-δ antibodies (Supplementary Fig. 10). This standard curve was then used to estimate the fraction of PKC-δ^{+} cells that would have been detectable by direct immunofluorescence labelling for GFP, on the basis of the quantification of native YFP fluorescence. In the behavioural experiments in which wild-type and PKC-δ::GluClα- CFP-iCre transgenic animals were classified on the basis of their infection rates (Table 1, Supplementary Table 5 and Supplementary Figs 5 and 7), this estimation method was applied to all experimental groups, including wild-type animals.

Electrophysiological slice recordings. Standard mouse brain slice preparation and whole-cell recording were performed as described[14]. In brief, coronal sections 250 µm thick were prepared with a Vibratome (VT1000S; Leica), using ice-cold glycerol-based artificial cerebrospinal fluid (composition in mM: 252 glycerol, 1.6 KCl, 1.2 NaH$_2$PO$_4$, 1.2 MgCl$_2$, 2.4 CaCl$_2$, 18 NaHCO$_3$, 11 glucose, oxygenated with 95% O$_2$/5% CO$_2$). Slices were allowed to recover for at least 1 h at 32 °C and kept at room temperature in standard artificial cerebrospinal fluid (ACSF; composition in mM: 126 NaCl, 1.6 KCl, 1.2 NaH$_2$PO$_4$, 1.2 MgCl$_2$, 2.4 CaCl$_2$, 18 NaHCO$_3$, 11 glucose, oxygenated with 95% O$_2$/5% CO$_2$). Cells expressing CFP or YFP were detected by infrared differential interference contrast and fluorescence video microscopy (Olympus BX51). Whole-cell voltage and current recordings were performed at 30 °C with a MultiClamp 700B amplifier and Digidata 1440A (Molecular Devices). The patch pipette, with a resistance of 5–8 MΩ, was filled with an intracellular solution containing (in mM) 135 potassium gluconate, 5 EGTA, 0.5 CaCl$_2$, 2 MgCl$_2$, 10 HEPES, 2 MgATP and 0.1 GTP, pH 7.2, 280–300 mosM. Data were sampled at 10 kHz, filtered at 3 kHz and analysed with pCLAMP 10 software.

Electrophysiological characterization of PKC-δ^{+} cells. The relation of genetic marker and electrophysiological type was analysed with Fisher's exact test with the null hypothesis that PKC-δ^{+} cells have the same proportion of late-firing neurons as do CEl neurons as a whole. We recorded 38 CEl neurons without knowing their genetic subtypes, 21 of which were late-firing neurons; 14 of the 17 non-late-firing neurons were regular spiking cells, and 3 were low-threshold bursting cells (Supplementary Tables 2 and 3). We also recorded 57 PKC-δ^{+} neurons on the basis of their CFP expression, and found 49 of them to be late-firing neurons and 8 of them to be non-late-firing neurons (all eight of these cells were regular spiking cells; Supplementary Table 4). Fisher's exact test revealed

that PKC-δ^+ neurons have a relatively homogeneous electrophysiological property of late-firing neurons in comparison with CEl neurons as a whole.

Stereotaxic surgery. Male mice 2–4 months old were deeply anaesthetized with either ketamine/xylazine or isoflurane, injected intraperitoneally with 500 μl of 20% mannitol (Phoenix Pharmaceuticals) and placed in a stereotaxic frame (Kopf). The skull was exposed and perforated with a stereotaxic mounted drill at the desired coordinates (Supplementary Table 6). For postoperative care, mice were injected intraperitoneally with ketoprofen (2 mg per kg body weight) and supplied for 10 days with drinking water containing 80 mg l^{-1} trimethoprim, 400 mg l^{-1} sulphamethoxazole and 200 mg l^{-1} ibuprofen.

Viral injections. Viral particles were delivered unilaterally by stereotaxic (see above) injection with through stereotaxic mounted motor-driven 32-gauge 45° bevelled steel cannulas (Micro 4 controller equipped with Hamilton system; World Precision Instruments) or glass capillaries (Micro4 controller equipped with Microject system; World Precision Instruments) to one to four injection sites per hemisphere at flow rates of 50–100 nl min^{-1}.

Pharmacogenetic silencing *in vitro*. Mice were unilaterally injected with a total of 2 μl of AAV GluClα or GluClβ virus (serotype 2; (1–10) × 10^{12} particles ml^{-1} in PBS) into the central amygdala (see above). Four weeks after surgery, slices (see above) were perfused for 20 min with 20 nM ivermectin in ACSF to induce pharmacogenetic silencing.

Pharmacogenetic silencing *in vivo*. Mice were injected bilaterally with a total of 2 μl of AAV GluClα or GluClβ virus (serotypes 2, 5 and 8; (1–10) × 10^{12} particles ml^{-1} in PBS) into the central amygdala (see above). Four weeks after surgery, on day 1, animals were habituated for 20 min to context A and given a single intraperitoneal dose of 1% ivermectin injectable solution (Phoenix) at 10 mg per kg^{-1} body weight, or vehicle. On day 2, animals were fear-conditioned in context A connected to a shock scrambler (Coulbourn Scientific or Med Associates fear conditioning systems were used). After 20 min, mice were exposed to training trials of tone (85 dBA, 2 kHz) for 20 s, immediately followed by a foot shock (0.5 mA) for 2 s, with an inter-trial interval of 198 s. To avoid masking of the consequences of the experimental manipulation by ceiling and floor effects, the number of training trials (three to six) was adjusted such that freezing levels averaged about 50% in the genetic/viral control. During the training session, behaviour was recorded with infrared sensors (Coulbourn fear conditioning system) or video cameras (Med Associates fear conditioning system). On day 3, mice were placed in context B, differentiated from context A by a modified wall, floor and ceiling. Before each trial, all surfaces were cleaned with water and disinfectant. After a 5-min baseline period, two test trials of a 30-s CS followed by a 1-min post-CS period were presented. Freezing was recorded with video cameras and scored either manually by an observer blind to the experimental group of the animal (Coulbourn fear conditioning system) or electronically (Med Associates fear conditioning system). Freezing during the CS and post-CS periods are represented as bins of these two test trials. Baseline freezing was averaged over 3 min before the first CS presentation.

To investigate the relationship between freezing and the extent of viral infection (Supplementary Fig. 5a–c), freezing data were correlated with the fraction of PKC-δ^+ expressing GluClβ, estimated as described above with native fluorescence for YFP and the standard curve in Supplementary Fig. 10.

To investigate differences in freezing between experimental and control groups, freezing in the experimental group was compared with genetic/viral control groups by a two-way ANOVA with infection rate as the blocking variable (Supplementary Table 5 and Supplementary Fig. 5d–f).

Retrograde tracing with CTB. CTB (0.5% in PBS; List) was delivered by stereotaxic (see above) iontophoresis with a positive-pulsed current of 5 μA for 2 min. The animals were killed 1 week after injection, and the brains were processed for IHC.

Optogenetic circuit dissection. Mice were unilaterally injected with a total of 2 μl of Cre-dependent AAV ChR2 virus (serotypes 2 and 5; (1–10) × 10^{12} particles ml^{-1} in PBS) into the central amygdala (see above). Slices (see above) were stimulated with an optic fibre (200 μm core diameter; ThorLabs) coupled to a 473-nm laser (CrystaLaser) and mounted on a three-dimensional micromanipulator (MPC365; Sutter) with the fibre end positioned on the edge of CEl.

Cre-dependent monosynaptic tracing with rabies virus. Cre-expressing cells were primed for subsequent infection and monosynaptic retrograde spread of EnvA-pseudotyped, glycoprotein gene-deleted rabies virus, by stereotaxic injection (see above) of 180 nl of Cre-dependent AAV expressing the avian receptor protein TVA and rabies B19 glycoprotein (AAV9-pEF1α-FLEX-GTB) into the central amygdala. The genomic structure of AAV9-pEF1α-FLEX-GTB is L_ITR-EF1α Pro-Kozak-(FLEX cassette[48] containing GFP-T2A-hTVA-E2A-hB19G)-STOP-WPRE-polyA-R_ITR. The start ATG has been deleted from all three genes, and transcription start is mediated by a Kozak sequence that precedes the FLEX cassette[49]. Genes were linked together through the use of 2A elements, which permit the expression of multiple genes under the control of a single promoter[50,51]. Both the TVA and B19G genes were codon-optimized for expression in mammalian cells. The virus was prepared through a crude lysate extraction of transfected 293T cells, and had a genomic titre of 10^8 particles ml^{-1}. Although reverse-complemented GFP is detectable in the AAV genome, Cre-expressing cells expressed undetectable levels of GFP, either through native fluorescence or antibody-amplified imaging, whereas TVA and B19G appeared to express normally.

Three weeks later, mice were injected at the same site with 180 nl of glycoprotein gene-deleted rabies virus[52] that had been pseudotyped with the avian sarcoma leucosis virus glycoprotein EnvA[40]. The glycoprotein gene had been replaced with mCherry. The resulting virus, (EnvA)SAD-dG-mCherry, was incapable of infecting mammalian neurons in the absence of its binding partner, TVA, and could not spread retrogradely in the absence of the rabies B19 glycoprotein[39]. The biological titre of 10^9 particles ml^{-1} was determined through infection of TVA-expressing 293T cells. One week after the second injection, the animals were killed and the brains were processed for IHC.

Pharmacogenetic silencing with *in vivo* electrophysiological recordings. Mice were injected bilaterally with a total of 2 μl of AAV GluClα virus (serotype 2; (1–10) × 10^{12} particles ml^{-1} in PBS) into the central amygdala (see above). Four weeks later, on day 1, mice were habituated to four CS presentations (total CS duration 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency 7.5 kHz, 80 dB sound pressure level) in context B. On day 2, mice were conditioned with five CS/unconditioned stimulus (0.6 mA, 1 s) pairings with inter-trial intervals of 20–180 s in context A (Coulbourn fear conditioning system). On days 3, 6 and 8, mice were exposed to four CS presentations in context B. During these periods, individual neurons were recorded extracellularly in freely behaving mice. Spikes of individual neurons were sorted by time–amplitude window discrimination and template matching as described previously[53,54]. Cluster quality was verified by quantifying the cluster separation[54] (Supplementary Fig. 8). Unit isolation was verified by using autocorrelation and cross-correlation histograms. Spike rasters and histograms were constructed by aligning sweeps relative to the CS onset, and CS-evoked responses were normalized to baseline activity using a Z-score transformation. Detailed methodological information about *in vivo* single-unit recordings of central amygdala neurons is presented in ref. 16.

47. Gong, S., Yang, X. W., Li, C. & Heintz, N. Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kγ origin of replication. *Genome Res.* **12,** 1992–1998 (2002).
48. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28,** 7025–7030 (2008).
49. Seidler, B. *et al.* A Cre-loxP-based mouse model for conditional somatic gene expression and knockdown *in vivo* by using avian retroviral vectors. *Proc. Natl Acad. Sci. USA* **105,** 10137–10142 (2008).
50. Furler, S., Paterna, J. C., Weibel, M. & Bueler, H. Recombinant AAV vectors containing the foot and mouth disease virus 2A sequence confer efficient bicistronic gene expression in cultured cells and rat substantia nigra neurons. *Gene Ther.* **8,** 864–873 (2001).
51. Szymczak, A. L. *et al.* Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnol.* **22,** 589–594 (2004).
52. Etessami, R. *et al.* Spread and pathogenic characteristics of a G-deficient rabies virus recombinant: an *in vitro* and *in vivo* study. *J. Gen. Virol.* **81,** 2147–2153 (2000).
53. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454,** 600–606 (2008).
54. Nicolelis, M. A. *et al.* Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl Acad. Sci. USA* **100,** 11041–11046 (2003).

ARTICLE

Encoding of conditioned fear in central amygdala inhibitory circuits

Stephane Ciocchi[1]*, Cyril Herry[1]*†, François Grenier[1], Steffen B. E. Wolff[1], Johannes J. Letzkus[1], Ioannis Vlachos[2], Ingrid Ehrlich[1]†, Rolf Sprengel[3], Karl Deisseroth[4], Michael B. Stadler[1], Christian Müller[1] & Andreas Lüthi[1]

The central amygdala (CEA), a nucleus predominantly composed of GABAergic inhibitory neurons, is essential for fear conditioning. How the acquisition and expression of conditioned fear are encoded within CEA inhibitory circuits is not understood. Using *in vivo* electrophysiological, optogenetic and pharmacological approaches in mice, we show that neuronal activity in the lateral subdivision of the central amygdala (CEl) is required for fear acquisition, whereas conditioned fear responses are driven by output neurons in the medial subdivision (CEm). Functional circuit analysis revealed that inhibitory CEA microcircuits are highly organized and that cell-type-specific plasticity of phasic and tonic activity in the CEl to CEm pathway may gate fear expression and regulate fear generalization. Our results define the functional architecture of CEA microcircuits and their role in the acquisition and regulation of conditioned fear behaviour.

The amygdala is a key brain structure involved in the acquisition and expression of conditioned fear responses[1–3]. In the classical circuit model of fear conditioning, the lateral nucleus of the amygdala is thought of as the primary site where associations between the conditioned stimulus (CS) and the unconditioned stimulus (US) are formed and stored[1–5]. In contrast to the lateral nucleus of the amygdala (LA), the central nucleus of the amygdala (CEA) has been considered to be primarily involved in the behavioural expression of conditioned fear responses. CEA output neurons, most of which are located in its medial subdivision (CEm), project to downstream targets in the brainstem and in the hypothalamus where they orchestrate conditioned autonomic and motor responses[6–8]. It is still unclear whether conditioned fear responses are triggered by activation or inhibition of these output neurons. Experiments involving lesions, pharmacological manipulations and electrical stimulation indicate that the activity of CEm output neurons drives conditioned fear responses[2]; however, the only *in vivo* recordings from identified CEA output neurons indicate the opposite[9].

Recent evidence also supports a role of the CEA in the learning process itself. Acute and reversible inactivation of the CEA during fear conditioning, or local blockade of NMDA (*N*-methyl-D-aspartate) receptors, result in impaired acquisition of conditioned fear responses[10,11]. This strongly indicates that activity-dependent plasticity within the CEA is necessary for the acquisition of fear conditioning[12,13]. Because CEm output neurons are under tight inhibitory control from the lateral and capsular subdivisions (together referred to as CEl)[14–17], a reduction in CEl to CEm inhibition might contribute to increased CEm output after fear conditioning. Although this scenario is indirectly supported by the observation that enhancing inhibitory activity in CEl by endogenous neuropeptides and exogenous substances, such as ethanol, has anxiolytic effects[18], the role of the intra-CEA inhibitory circuitry in the acquisition and expression of conditioned fear responses is not known.

Differential role of CEl and CEm

To test the impact of neuronal activity in CEm on freezing behaviour, we selectively activated or inhibited CEm neurons using optogenetic and pharmacological approaches. Activation of CEm neurons was achieved by bilateral targeted injection of a virus expressing channelrhodopsin-2 (ChR2) in neurons[19,20] (Fig. 1a, b and Supplementary Fig. 1). In freely moving animals, bilateral activation of CEm induced strong and reversible freezing responses (Fig. 1c).

Conversely, we used microiontophoresis of a fluorescently labelled GABA$_A$ receptor agonist[21] (muscimol-bodipy (BPY)) to inhibit neuronal activity in a targeted and reversible manner (Fig. 1d). Bilateral inactivation of CEm, or of the entire CEA (CEm and CEl), did not elicit freezing behaviour (Fig. 1e). In contrast, inactivation of CEl alone induced unconditioned freezing (Fig. 1e). These results indicate that neuronal activity in CEm is necessary and sufficient for driving freezing behaviour, and that CEm is under tonic inhibitory control from CEl.

We next examined the contribution of distinct CEA subnuclei to the acquisition of conditioned freezing. Mice were trained in a discriminative fear conditioning paradigm (see Methods). Twenty-four hours after conditioning, mice exhibited freezing behaviour upon presentation of the conditioned stimulus (CS$^+$; $P < 0.001$ versus pre-conditioning baseline). Consistent with a previous study[11], targeted bilateral inactivation of the entire CEA during fear conditioning resulted in a profound memory deficit when measured 24 h later in the absence of muscimol-BPY (Fig. 1f). A comparable memory deficit was observed after inactivation of CEl, but not of CEm (Fig. 1f). Notably, inactivation of CEA or CEl did not impair US detection, and after wash-out of muscimol-BPY, animals could be fear conditioned, indicating that muscimol-BPY did not cause irreversible damage (Supplementary Fig. 2).

We next examined the role of CEl and CEm for memory retrieval or expression by local application of muscimol-BPY 24 h after conditioning (Fig. 1g). In contrast to the acquisition phase, we found that whereas inactivation of the entire CEA or CEm resulted in a retrieval/expression deficit, inactivation of CEl did not reduce conditioned freezing levels, although occlusion by unconditioned freezing is possible (Fig. 1g and Supplementary Fig. 3). Together, these data indicate a functional dissociation of CEl and CEm during the acquisition and expression of conditioned fear responses, and indicate a role for activity-dependent neuronal plasticity in CEl in acquisition.

Figure 1 | Differential role of CEl and CEm in fear expression and acquisition. a, Coronal section of the mouse brain indicating the location of the central amygdala (CEA). CEl/CEm, lateral/medial subdivisions of CEA. Numbers indicate the antero-posterior coordinates caudal to bregma. **b**, Red fluorescent neurons in CEm infected with AAV-ChR22A-tdimer. Scale bar, 100 μm. **c**, Left: example experiment illustrating rapid and reversible freezing induced by bilateral stimulation of ChR2-expressing CEm neurons with 10 s of blue light (inter-stimulation intervals, 30–60 s). Right: summary data demonstrating significant light-induced freezing responses in AAV-ChR22A-tdimer infected animals, but not in sham-operated controls. **d**, Epifluorescence image illustrating microiontophoretic application of fluorescently labelled muscimol (muscimol-BPY) targeted at CEm, CEl, or the entire CEA. **e**, Top: experimental protocol. Bottom: inactivation of CEl induced unconditioned freezing. Inactivation of CEm, or the entire CEA, had no effect on freezing. Control mice received BPY only. **f**, Top: muscimol-BPY was applied during fear conditioning. Animals were tested drug-free 24 h later. Bottom: compared with animals injected with BPY only, inactivation of CEl or CEA, but not of CEm, prevented fear acquisition. **g**, Top: animals were fear conditioned in the absence of muscimol-BPY and tested 24 h later. Muscimol-BPY was applied before animals were re-tested on the same day. Bottom left: at test, all experimental groups exhibited equal freezing levels before muscimol application. Bottom right: compared with animals injected with BPY only, inactivation of CEm or CEA, but not CEl, impaired fear expression. All error bars indicate mean ± s.e.m. *$P < 0.05$, **$P < 0.01$. Statistical analysis is shown in the Supplementary Information.

Organization of CEA inhibitory networks

Next, we investigated fear-conditioning-induced changes in CS-evoked neuronal firing in the CEl. Mice were implanted with chronic recording electrodes and 167 units located in the CEl were recorded (Supplementary Fig. 4). Two classes of CEl units exhibiting opposite changes in CS-evoked activity after fear conditioning were revealed by comparing z-scored CS responses 24 h after conditioning to baseline levels measured during habituation. Whereas 30% of units acquired an excitatory response (CEl_{on} neurons) (Fig. 2a), 25% of CEl neurons displayed a strong inhibitory response to the CS^+ after fear conditioning (CEl_{off} neurons) (Fig. 2b). The rest of the units (45%) did not exhibit any tone-evoked responses. Changes in CS^+-evoked responses were already detectable during fear conditioning, and in animals exhibiting behavioural discrimination, both CEl_{on} and CEl_{off} units exhibited discriminating neuronal responses (Supplementary Fig. 4). Thus, fear conditioning induces rapid, specific and persistent changes in CS-evoked activity of CEl neurons.

The inverse direction of fear-conditioning-induced plasticity in CEl_{on} and CEl_{off} neurons indicated the possibility that inhibitory responses of CEl_{off} neurons were mediated by local inputs from CEl_{on} neurons. Latency analysis of CS responses in CEl_{on} and CEl_{off} neurons revealed that CS-evoked excitation in CEl_{on} neurons started before CEl_{off} neurons were inhibited (Supplementary Fig. 4). The short onset latency of CS-evoked excitation in CEl_{on} neurons

(<15 ms) indicates that they may, like CEm neurons[22,23], receive direct input from sensory thalamus[24]. Cross-correlating spontaneously occurring spikes of simultaneously recorded CEl_{on} and CEl_{off} neurons revealed substantial, yet asymmetrical, short-latency inhibitory interactions between the two classes of neurons (Fig. 2c, f; CEl_{on} to CEl_{off}, 9 of 35 pairs; CEl_{off} to CEl_{on}, 3 of 35 pairs; $P < 0.05$, binomial test). Inhibitory cross-correlations among CEl_{on} neurons (0 of 22 possible connections) or among CEl_{off} neurons (2 of 54 possible connections) were rare. Thus, fear conditioning leads to a shift in the balance of activity between distinct functional classes of CEl neurons embedded into highly organized local inhibitory circuits.

On the basis of previous anatomical and *in vitro* electrophysiological studies in rats describing an inhibitory GABAergic projection from CEl to CEm[13–17], we examined anatomical and functional connectivity between CEl and CEm. First, we locally injected a retrogradely tracing virus (herpes simplex virus 1 (HSV-1))[25] into CEl or CEm. Whereas injections into CEm resulted in intense retrograde labelling of neurons in CEl, CEm remained largely devoid of GFP after injections into CEl (Supplementary Fig. 5), indicating that CEl projects to CEm in a mostly unidirectional manner[26]. Next, to address whether identified CEl_{on} or CEl_{off} neurons project to CEm, we performed intracellular recordings in anaesthetized animals which were previously fear conditioned. Like in awake and behaving animals, CEl neurons recorded in anaesthetized mice were spontaneously active (Fig. 2d, e) and displayed both excitatory (5 of 12 neurons) and inhibitory (2 of 12 neurons) CS responses (Fig. 2d, e). Morphological reconstruction of neurobiotin-filled neurons revealed that the axons of both subtypes arborize locally within CEl, and send collaterals to CEm (Fig. 2d, e). To test whether CEl_{on} and CEl_{off} neurons functionally inhibit CEm neurons, we performed simultaneous multi-site single-unit recordings in CEl and CEm and cross-correlated spiking activity between identified pairs of neurons. Both CEl_{on} and CEl_{off} neurons exhibited inhibitory interactions with CEm neurons (Supplementary Fig. 6). No interactions in the reverse direction (from CEm to CEl) were found. These findings provide strong evidence that two distinct subclasses of CEl neurons inhibit CEm neurons *in vivo* (Fig. 2f). Considering that CEl_{on} and CEl_{off} neurons exhibited opposite changes in CS-evoked firing during fear conditioning, this raises the question of whether at the level of CEm output neurons fear conditioning results in CS-evoked inhibition or disinhibition.

To address this, we identified CEm output neurons by placing a stimulation electrode in the mesencephalic axon bundle, a fibre tract containing CEm projections to brainstem targets[9]. In four out of six cases, reliable time-locked antidromic responses could be evoked under anaesthesia after identification of neuronal responses in CEm units in freely moving fear-conditioned mice (Fig. 3a, b). Consistent with the finding that unconditioned freezing can be induced by activation of CEm or by inhibition of CEl, 83% of neurons located in CEm exhibited a marked increase in CS^+-evoked firing 24 h after fear conditioning (Fig. 3c), including all four identified brainstem-projecting CEm cells. The remainder of the units did not exhibit any CS^+-evoked responses. Like in CEl, CS responses of CEm output neurons started to change during fear conditioning and discriminated between CS^+ and CS^- (unpaired control stimulus) presentations (Supplementary Fig. 7). CEm neurons displayed a bi-phasic CS response (Supplementary Fig. 7). The first component was brief with a short (10–15 ms) onset latency similar to excitatory CS responses of CEl_{on} neurons. The second component was more sustained, paralleling the time course of inhibitory CEl_{off} responses, albeit with a slightly longer onset latency (Supplementary Fig. 7). These findings are consistent with the notion that conditioned CS responses of CEm output neurons reflect the integration of both excitatory and disinhibitory inputs (see circuit scheme in Supplementary Fig. 8).

Tonic inhibition and fear generalization

Given that CEm output is under tonic inhibitory control from CEl, this raises the question as to whether spontaneous activity in CEl and
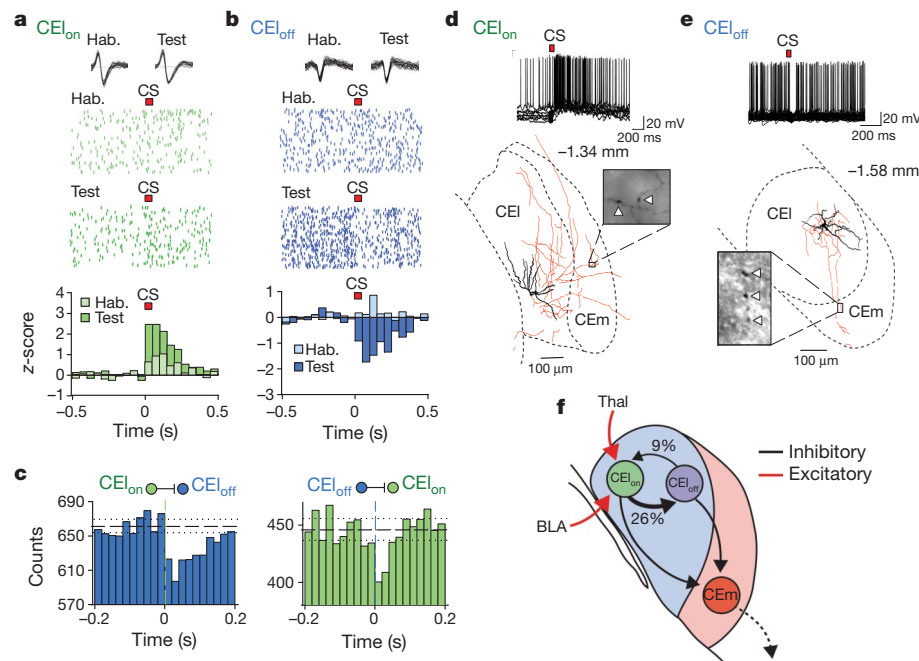
Figure 2 | Fear conditioning induces cell-type-specific plasticity in CEl inhibitory circuits. a, b, A subpopulation of CEl units (CEl$_{on}$ neurons) exhibited an increase in CS-evoked firing after fear conditioning (**a**), whereas another subpopulation (CEl$_{off}$ neurons) acquired an inhibitory CS response (**b**). Example raster plots (top) and normalized and averaged population peri-stimulus time histograms (PSTHs; bottom) from all CEl$_{on}$ or CEl$_{off}$ units, respectively, are shown, illustrating fear conditioning-induced changes in CS-evoked firing. The duration of the auditory stimulus is indicated (red bar; CS). Traces show superimposed spike waveforms recorded during habituation and test, respectively. **c,** Asymmetric functional organization of local inhibitory circuits in CEl. Left panel: averaged cross-correlation analysis ($n = 9$ pairs) illustrating a short latency inhibitory interaction between CEl$_{on}$ and CEl$_{off}$ neurons in 26% of all recorded pairs. Dashed vertical line indicates time of reference spikes from CEl$_{on}$ units. Right panel: averaged cross-correlation

analysis ($n = 3$ pairs) illustrating a short latency inhibitory interaction between CEl$_{off}$ and CEl$_{on}$ neurons in 9% of all recorded pairs. Dashed horizontal line indicates mean; dotted lines indicate 95% confidence interval. **d, e,** CEl$_{on}$ and CEl$_{off}$ neurons send axon collaterals to CEm. Top: example intracellular recordings of a CEl$_{on}$ and a CEl$_{off}$ neuron spontaneously active at resting membrane potential (V_m). CS presentation induced an increase (in CEl$_{on}$ neurons) or a decrease (in CEl$_{off}$ neurons) in firing. Bottom: reconstruction of the neurobiotin-filled neurons revealed extensive local axonal branching and projections targeting CEm. Soma and dendrites are indicated in black, axon is shown in red. Inset shows putative synaptic contacts as suggested by the presence of axonal boutons in CEm (arrowheads). Numbers indicate the antero-posterior coordinates caudal to bregma. **f,** Schematic illustrating the organization of CEA inhibitory circuits based on electrophysiological and morphological data. Statistical analysis is shown in the Supplementary Information. Thal., Thalamus.

CEm is subject to regulation, and how plasticity of spontaneous activity might contribute to the encoding of conditioned fear responses. Indeed, spontaneous activity of CEm output neurons was markedly decreased after fear conditioning (Fig. 4a). Conversely, CEl$_{off}$ neurons exhibited increased spontaneous activity after fear conditioning, whereas on average CEl$_{on}$ neurons showed a slight decrease (Fig. 4a).

Phasic z-scored CS-evoked neuronal activity was highly correlated with freezing behaviour in all three neuronal subpopulations (Fig. 4b and Supplementary Fig. 9). However, because z-scores reflect CS-evoked neuronal activity normalized to pre-CS tonic activity, an increase in the z-score could reflect a net increase in the phasic (for example, synaptic) CS response, or a decrease in the absolute level or in the variability of pre-CS tonic activity. Plotting changes in tonic versus changes in phasic activity revealed that the two phenomena were correlated (Supplementary Fig. 10). However, plasticity of phasic and tonic activity seem to be independent processes, as fear conditioning did not affect variability of tonic activity and also increased phasic CS responses in neurons that did not exhibit concomitant changes in tonic activity (Supplementary Fig. 10).

What might be the behavioural relevance of plasticity of tonic activity in CEA inhibitory circuits? After fear conditioning, absolute and z-scored levels of tonic activity were only poorly correlated with freezing (Fig. 4c and Supplementary Fig. 11). Fear-conditioning-induced changes in tonic activity were not limited to periods of CS$^+$ exposure, but were also manifest during CS$^-$ stimulation (Supplementary Fig. 12), indicating that plasticity of tonic activity could regulate the signal-to-noise ratio by linearly offsetting both CS$^+$- and CS$^-$-evoked phasic responses. This would be expected to affect fear generalization. Consistent with this notion, changes in tonic activity

predicted CS$^+$ versus CS$^-$ discrimination at the behavioural level (Fig. 4d). Specifically, a decrease in tonic activity of CEm output neurons was associated with generalization, whereas CEl$_{off}$ neurons and CEl$_{on}$ neurons exhibited the inverse correlation (Fig. 4d and Supplementary Fig. 8), consistent with tonic inhibition of CEm output neurons by both CEl$_{on}$ and CEl$_{off}$ neurons. In addition, a receiver operating characteristic analysis (ROC) of the pooled CEl population revealed that changes in tonic activity of CEl neurons were significantly higher ($P < 0.01$) in generalizing mice compared to discriminating ones. Notably, changes in tonic activity were already present before CS onset (Supplementary Fig. 13), indicating that generalization is associated with a different functional network state, and that the stimulus specificity of conditioned fear responses is regulated by concerted changes in tonic and phasic activity within the neuronal circuitry of the CEA.

Discussion

Using targeted and reversible pharmacological and optogenetic approaches, we show that conditioned and unconditioned freezing behaviour is driven by CEm output neurons which are under tonic inhibitory control originating in CEl. Moreover, our study identifies CEl as an essential component of the neuronal circuitry underlying the acquisition of conditioned fear. We found that CEl contains two functionally distinct subpopulations of neurons forming highly organized local inhibitory circuits which inhibit CEm output neurons. Notably, an accompanying study[27] shows that at least one of these functionally defined subpopulations of CEl neurons (CEl$_{off}$ neurons) largely overlaps with a genetically defined neuronal subtype (PKC-δ^+ neurons). Our data indicate that whereas conditioned fear responses are driven
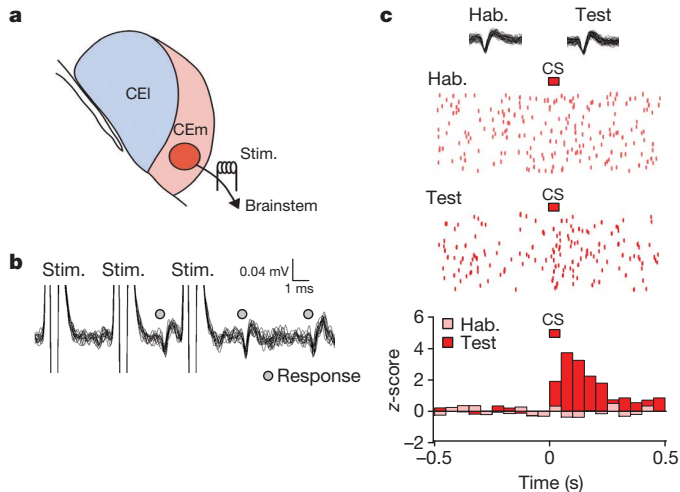
Figure 3 | Fear conditioning induces disinhibition of CEm output neurons. **a**, Identification of CEm output neurons by stimulation of the mesencephalic axon bundle. **b**, Stimulation of the mesencephalic axon bundle elicits antidromic spikes which exhibited low temporal jitter (<0.1 ms), and followed high-frequency (200 Hz) stimulation. **c**, CEm units exhibited an increase in CS-evoked firing after fear conditioning. Example raster plots (top) and normalized and averaged population PSTHs (bottom) from all CEm units illustrating fear-conditioning-induced changes in CS-evoked firing. The duration of the auditory stimulus is indicated (red bar; CS). Traces show superimposed spike waveforms recorded during habituation and test, respectively. Statistical analysis is shown in Supplementary Information.

by CS-evoked disinhibition of CEm output neurons, cell-type-specific plasticity of tonic inhibitory network activity within the CEl/CEm circuitry regulates generalization of conditioned fear responses.

Inactivation of CEl during the acquisition of fear conditioning interfered with learning. This suggests a role for synaptic plasticity at glutamatergic inputs onto CEl_{on} neurons during fear conditioning. CEl receives glutamatergic input from various brain structures including the basolateral complex of the amygdala (BLA), insular cortex, and brainstem, in particular from the parabrachial nucleus[12]. Afferents originating from the parabrachial nucleus make very strong and reliable synapses onto CEl neurons[28], and might function as a teaching signal enabling the induction of synaptic plasticity at other inputs. However, although long-term potentiation can be induced at various inputs to the CEl in slice preparations[29,30], the role of specific afferent pathways in fear conditioning remains to be determined.

After fear conditioning, CEm output neurons exhibited CS-evoked bi-phasic excitation, consisting of a brief short-latency response followed by a slower second component. Considering the very short onset latency of the first component (less than 15 ms) it is likely to be driven by direct excitatory input from auditory thalamic nuclei[12,22,23,31]. Because CEl_{on} neurons also exhibited short-latency excitatory responses, thalamo-CEm excitation may be terminated by feedforward inhibition through the CEl_{on} pathway. Consistent with this scenario, sparse inputs from auditory thalamus (suprageniculate and posterior intralaminar nuclei) to the capsular part of CEl have been described[24]. Alternatively, termination of short-latency responses might also reflect feedforward inhibition mediated by intercalated cell clusters (ITCs)[32,33]. The observation that short-latency excitation of CEm output neurons increased with fear conditioning indicates that thalamo-CEm synapses might be strengthened, possibly involving NMDA receptor-dependent long-term potentiation[30].

The second component of CS-evoked responses of CEm output neurons, which had a much longer duration and contained most of the spikes, most likely reflects disinhibitory input from CEl, and possibly from nearby ITCs[34] as well as direct excitatory input from BLA. Several arguments support a role for disinhibition via the CEl_{off} pathway. First,
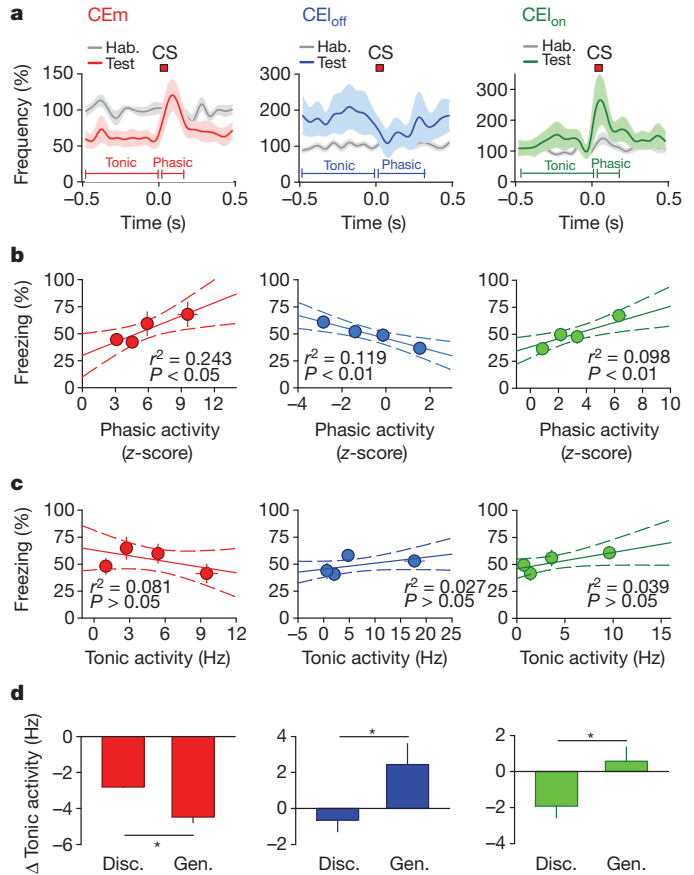
Figure 4 | Cell-type-specific plasticity of tonic activity regulates fear generalization. **a**, Normalized population PSTHs of CEm (left), CEl_{off} (middle) and CEl_{on} neurons (right). Fear conditioning was associated with decreased tonic activity in CEm and CEl_{on} neurons, and increased tonic activity in CEl_{off} neurons. Shaded area indicates s.e.m. **b**, Phasic z-scored CS responses of CEm (left), CEl_{off} (middle) and CEl_{on} neurons (right) are correlated with freezing behaviour. Correlations were obtained by plotting the averaged z-scored responses for all neurons of a given subtype for each animal over blocks of two CSs. Plots include both CS^+ and CS^- responses. Data were averaged and binned (bins contain an equal number of data points). Linear correlations were performed using the non-binned raw data (Supplementary Fig. 9). P values indicate significance levels of Pearson's correlation coefficients. Dashed lines indicate 95% confidence intervals. All error bars indicate mean ± s.e.m. **c**, Tonic activity does not correlate with freezing behaviour. **d**, Cell-type-specific plasticity of tonic activity predicts fear generalization. Changes in tonic activity were different when animals exhibited CS^+/CS^- discrimination (Disc.; freezing ratio $CS^+/CS^- > 1.4$) or when animals generalized (Gen.; freezing ratio $CS^+/CS^- < 1.4$). Whereas CEm units showed a stronger decrease in tonic activity upon generalization, CEl_{off} neurons and CEl_{on} neurons exhibited the inverse correlation. $*P < 0.05$. Statistical analysis is shown in the Supplementary Information.

CEl_{off} neurons can project to CEm, functionally inhibit CEm output neurons and regulate conditioned freezing behaviour (see also the accompanying paper[27]). Second, pharmacological inactivation of CEl induced CEm-dependent freezing behaviour, demonstrating that CEm is under tonic inhibitory control from CEl. Third, CS-evoked inhibition of CEl_{off} neurons started right before the onset of slow excitation in CEm. Moreover, CS responses of CEl_{off} and CEm units exhibited a similar time course, which was much slower than excitatory CS responses of BLA neurons under comparable conditions[21]. Finally, pharmacological inactivation of CEl during fear conditioning resulted in a learning deficit, indicating that activity-dependent plasticity of CEl to CEm signalling is necessary for the acquisition of conditioned fear responses. Together, these findings strongly indicate that disinhibition

through the CEl$_{off}$ pathway contributes to fear-conditioning-induced changes in CEm output.

Fear conditioning induced cell-type-specific plasticity of tonic activity. Strong decreases in tonic activity of CEm neurons, or increases in tonic activity of CEl$_{off}$ neurons, predicted generalization of behavioural responses to the CS$^-$. Previous studies have implicated auditory cortex or cortico-LA connections in stimulus discrimination and fear generalization[35–37]. However, because the CEA is downstream of auditory cortex and the LA, regulation of fear generalization in the CEA might override stimulus discrimination established in these upstream structures. This would enable animals to re-adjust the appropriate degree of fear generalization according to internal state and environmental demands.

Our results are consistent with studies on appetitive conditioning paradigms[38,39] indicating that CEA can process fear-related information in series with the BLA, or independently, in a parallel manner. In a serial processing mode, inhibitory circuits in the CEA may further select instructive signals sent out by the BLA, thereby increasing the computational power and the possibilities for modulation of the amygdala circuitry. In addition, the CEA might function in parallel with, or even independently of, BLA and directly elicit conditioned or unconditioned emotional behaviours in response to specific internal states associated with altered neuromodulatory input, such as chronic pain[40]. Interestingly, CEA output has also been suggested to exert strong control over basal forebrain cholinergic circuits[41,42], indicating that the CEA also affects processing in higher brain structures like the neocortex.

Our data reveal that inhibitory circuits in the CEA are highly organized, and establish important, but distinct, roles for plasticity of phasic and tonic inhibitory network activity in fear conditioning. Inhibitory circuits, such as those in striatum, have been proposed to be particularly effective in output selection based on winner-share-all mechanisms[43]. Thus, CEA inhibitory circuits may fine-tune and select output pathways targeting distinct downstream structures depending on the pattern of afferent synaptic input and local neuromodulatory activity. Given that CEA circuitry is thought to be organized similarly to striatal circuits[15], this may indicate that coordinated changes in phasic and tonic inhibition are a widespread mechanism regulating stimulus specificity of associative learning in the CNS.

METHODS SUMMARY

Behaviour and pharmacological inactivations. Mice were submitted to a discriminative auditory fear conditioning paradigm in which the CS$^+$, but not the CS$^-$, was paired with an US (mild foot shock). Freezing behaviour was quantified using an automatic infrared beam detection system as previously described[21]. Bilateral inactivation of the CEA or CEA subdivisions was achieved using micro-iontophoretic injection of fluorescently labelled muscimol before fear conditioning or retrieval test. Behavioural discrimination was analysed using clustering and ROC analyses as described (see Methods).

Electrophysiological recordings and analysis. Individual neurons were recorded extracellularly in freely behaving mice. Spikes of individual neurons were sorted by time-amplitude window discrimination and template matching as previously described[21,44]. Cluster quality was verified by quantifying the cluster separation[44] (Supplementary Fig. 14). Unit isolation was verified using auto- and cross-correlation histograms. Spike rasters and histograms were constructed by aligning sweeps relative to the CS onset, and CS-evoked responses were normalized to baseline activity using a z-score transformation.

Virus injections and optical stimulation. Animals were stereotaxically injected with an AAV serotype 2/7 expressing ChR22A-tdimer. Behavioural experiments were performed after 4 weeks of recovery and expression time and 3 days of handling. Optic fibres with a diameter of 200 μm were inserted into chronically implanted guide cannulae. ChR2-expressing cells were stimulated using a 473-nm laser. Freezing with and without light stimulation was quantified as described above.

Intracellular recordings and morphological reconstructions. Intracellular recordings were obtained from head-fixed animals under chloral hydrate anaesthesia (400 mg kg^{-1}) using standard methods. After completion of the recordings, animals were transcardially perfused and the brain kept for morphological reconstruction of the neurobiotin-filled recorded neurons.

1. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).
2. Davis, M. The role of the amygdala in conditioned and unconditioned fear and anxiety. In *The Amygdala* (ed., Aggleton, J. P.) 213–288 (Oxford Univ. Press, 2000).
3. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nature Rev. Neurosci.* **5,** 844–852 (2004).
4. Sigurdsson, T., Doyère, V., Cain, C. K. & LeDoux, J. E. Long-term potentiation in the amygdala: a cellular mechanism of fear learning and memory. *Neuropharmacology* **52,** 215–227 (2007).
5. Sah, P., Westbrook, R. F. & Lüthi, A. Fear conditioning and long-term potentiation: what really is the connection? *Ann. NY Acad. Sci.* **1129,** 88–95 (2008).
6. Krettek, J. E. & Price, J. L. A description of the amygdaloid complex in the rat and cat with observations on intra-amygdaloid axonal connections. *J. Comp. Neurol.* **178,** 255–279 (1978).
7. Veening, J. G., Swanson, L. W. & Sawchenko, P. E. The organization of projections from the central nucleus of the amygdala to brainstem sites involved in central autonomic regulation: a combined retrograde transport-immunohistochemical study. *Brain Res.* **303,** 337–357 (1984).
8. LeDoux, J. E., Iwata, J., Cicchetti, P. & Reis, D. J. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *J. Neurosci.* **8,** 2517–2529 (1988).
9. Pascoe, J. P. & Kapp, B. S. Electrophysiological characteristics of amygdaloid central nucleus neurons during Pavlovian fear conditioning in the rabbit. *Behav. Brain Res.* **16,** 117–133 (1985).
10. Goosens, K. A. & Maren, S. Pretraining NMDA receptor blockade in the basolateral complex, but not the central nucleus, of the amygdala prevents savings of conditional fear. *Behav. Neurosci.* **117,** 738–750 (2003).
11. Wilensky, A. E., Schafe, G. E., Kristensen, M. P. & LeDoux, J. E. Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J. Neurosci.* **26,** 12387–12396 (2006).
12. Samson, R. D., Duvarci, S. & Paré, D. Synaptic plasticity in the central nucleus of the amygdala. *Rev. Neurosci.* **16,** 287–302 (2005).
13. Ehrlich, I. *et al.* Amygdala inhibitory circuits and the control of fear memory. *Neuron* **62,** 757–771 (2009).
14. Sun, N., Yi, H. & Cassell, M. D. Evidence for a GABAergic interface between cortical afferents and brainstem projection neurons in the rat central extended amygdala. *J. Comp. Neurol.* **340,** 43–64 (1994).
15. Cassell, M. D., Freedman, L. J. & Shi, C. The intrinsic organization of the central extended amygdala. *Ann. NY Acad. Sci.* **877,** 217–241 (1999).
16. Veinante, P. & Freund-Mercier, M. J. Branching patterns of central amygdaloid nucleus afferents in the rat: Single axon reconstructions. *Ann. NY Acad. Sci.* **985,** 552–553 (2003).
17. Huber, D., Veinante, P. & Stoop, R. Vasopressin and oxytocin excite distinct neuronal populations in the central amygdala. *Science* **308,** 245–248 (2005).
18. Roberto, M., Madamba, S. G., Moore, S. D., Tallent, M. K. & Siggins, G. R. Ethanol increases GABAergic transmission at both pre- and postsynaptic sites in rat central amygdala neurons. *Proc. Natl Acad. Sci. USA* **100,** 2053–2058 (2003).
19. Gradinaru, V. *et al.* Targeting and readout strategies for fast optical neural control *in vitro* and *in vivo*. *J. Neurosci.* **27,** 14231–14238 (2007).
20. Tang, W. *et al.* Faithful expression of multiple proteins via 2A-peptide self-processing: A versatile and reliable method for manipulating brain circuits. *J. Neurosci.* **29,** 8621–8629 (2009).
21. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454,** 600–606 (2008).
22. LeDoux, J. E., Ruggiero, D. A. & Reis, D. J. Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat. *J. Comp. Neurol.* **242,** 182–213 (1985).
23. Turner, B. H. & Herkenham, M. Thalamoamygdaloid projections in the rat: a test of the amygdala's role in sensory processing. *J. Comp. Neurol.* **313,** 295–325 (1991).
24. Linke, R., Braune, G. & Schwegler, H. Differential projection of the posterior paralaminar thalamic nuclei to the amygdaloid complex in the rat. *Exp. Brain Res.* **134,** 520–532 (2000).
25. Lima, S. Q., Hromadka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS ONE* **4,** e6099 (2009).
26. Pitkänen, A., Savander, V. & LeDoux, J. E. Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends Neurosci.* **20,** 517–523 (1997).
27. Haubensak, W. *et al.* Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature* doi:10.1038/nature09553 (this issue).
28. Delaney, A. J., Crane, J. W. & Sah, P. Noradrenaline modulates transmission at a central synapse by a presynaptic mechanism. *Neuron* **56,** 880–892 (2007).
29. Fu, Y. & Shinnick-Gallagher, P. Two intra-amygdaloid pathways to the central amygdala exhibit different mechanisms of long-term potentiation. *J. Neurophysiol.* **93,** 3012–3015 (2005).
30. Lopez de Armentia, M. & Sah, P. Bidirectional synaptic plasticity at nociceptive afferents in the rat central amygdala. *J. Physiol. (Lond.)* **581,** 961–970 (2007).
31. Samson, R. D. & Paré, D. Activity-dependent synaptic plasticity in the central nucleus of the amygdala. *J. Neurosci.* **25,** 1847–1855 (2005).
32. Millhouse, O. E. The intercalated cells of the amygdala. *J. Comp. Neurol.* **247,** 246–271 (1986).

33. Paré, D. & Smith, Y. The intercalated cell masses project to the central and medial nuclei of the amygdala in cats. *Neuroscience* **57,** 1077–1090 (1993).
34. Paré, D., Quirk, G. J. & LeDoux, J. E. New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* **92,** 1–9 (2004).
35. Thompson, R. F. The role of the cerebral cortex in stimulus generalization. *J. Comp. Physiol. Psychol.* **55,** 279–287 (1962).
36. Jarrell, T. W., Gentile, C. G., Romanski, L. M., McCabe, P. M. & Schneiderman, N. Involvement of cortical and thalamic auditory regions in retention of differential bradycardia conditioning to acoustic conditioned stimuli in rabbits. *Brain Res.* **412,** 285–294 (1987).
37. Shaban, H. *et al.* Generalization of amygdala LTP and conditioned fear in the absence of presynaptic inhibition. *Nature Neurosci.* **9,** 1028–1035 (2006).
38. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26,** 321–352 (2002).
39. Balleine, B. W. & Killcross, S. Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* **29,** 272–279 (2006).
40. Neugebauer, V., Galhardo, V., Maione, S. & Mackey, S. C. Forebrain pain mechanisms. *Brain Res. Brain Res. Rev.* **60,** 226–242 (2009).
41. Jolkkonen, E., Miettinen, R., Pikkarainen, M. & Pitkänen, A. Projections from the amygdaloid complex to the magnocellular cholinergic basal forebrain in rat. *Neuroscience* **111,** 133–149 (2002).
42. Gozzi, A. *et al.* A neural switch for active and passive fear. *Neuron* **67,** 656–666 (2010).
43. Wickens, J. R., Arbuthnott, G. W. & Shindou, T. Simulation of GABA function in the basal ganglia: computational models of GABAergic mechanisms in basal ganglia function. *Prog. Brain Res.* **160,** 313–329 (2007).
44. Nicolelis, M. A. L. *et al.* Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl Acad. Sci. USA* **100,** 11041–11046 (2003).

Author Contributions S.C. and C.H. initiated the project. S.C., C.H., F.G., S.B.E.W. and C.M. performed the experiments. S.C., C.H., F.G., S.B.E.W., I.V., M.B.S. and A.L. analysed the data. K.D. and R.S. provided constructs and advice. S.C., C.H., F.G., S.B.E.W., I.E. and J.J.L. contributed to the experimental design and interpretation. A.L. conceived the project, contributed to the experimental design and interpretation, analysed data and wrote the manuscript. S.C. and C.H. contributed equally. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.L. (andreas.luthi@fmi.ch).

METHODS

Animals. Male C57BL6/J mice (2–3 months old; Harlan Ltd) were individually housed for 7 days before all experiments, under a 12 h light/dark cycle, and provided with food and water ad libitum. All animal procedures were executed in accordance with institutional guidelines and were approved by the Veterinary Department of the Canton of Basel-Stadt.

Behaviour. Fear conditioning and fear test took place in two different contexts (context A and B). The conditioning and test boxes and the floor were cleaned with 70% ethanol or 1% acetic acid before and after each session, respectively. To score freezing behaviour, an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments) was used. Mice were considered to be freezing if no movement was detected for 2 s and the measure was expressed as a percentage of time spent freezing. To ensure that our automatic system scores freezing rather than just immobility, we previously compared the values obtained with those measured using a classical time-sampling procedure during which an experimenter blind to the experimental conditions determined the mice to be freezing or not freezing every 2 s (defined as the complete absence of movement except for respiratory movements)[21]. The values obtained were 95% identical and the automatic detection system was therefore used throughout the experimental sessions. On day 1, mice were submitted to a habituation session in context B, in which they received 4 presentations of the CS^+ and the CS^- (total CS duration of 30 s, consisting of 50-ms pips repeated at 0.9 Hz, 2-ms rise and fall; pip frequency: 7.5 kHz or white noise, 80 dB sound pressure level). Discriminative fear conditioning was performed on day 2 by pairing the CS^+ with a US (1-s foot shock, 0.6 mA, 5 CS^+/US pairings; inter-trial interval: 20–180 s). The onset of the US coincided with the offset of the CS^+. The CS^- was presented after each CS^+/US association but was never reinforced (5 CS^- presentations, inter-trial interval: 20–180 s). The frequencies used for CS^+ and CS^- were counterbalanced across animals. On day 3, conditioned mice were submitted to fear retrieval in context B, during which they received four and four presentations of the CS^- and the CS^+, respectively. When plotting the evolution of the freezing response in relation to the CS presentation, we found that the freezing probability starts increasing around 1.3 s after the presentation of the first pip of the sequence (Supplementary Fig. 15). After the first two pip presentations, the probability of freezing has already sharply increased. These data are consistent with an equivalence of individual pips, justifying the analyses of electrophysiological responses at the level of individual pips. Pharmacological experiments were performed using a non-discriminative conditioning protocol.

US-induced flinching behaviour and vocalizations were compared in the presence and absence of muscimol in freely moving mice. Foot-shock amplitudes ranged from 0.1 to 1 mA (3-ms duration).

To classify the freezing behaviour we used an expectation-maximization algorithm that provided the maximum likelihood estimates for a gaussian mixture model (GMM). That is, given the two-dimensional behavioural data (CS^+ freezing, CS^- freezing) we estimated the probability density $f(x) \in F$ that is most likely to have generated the data. We assumed that the family F is a two-component mixture of Gaussian functions:

$$f(x; \theta) = \sum_{k=1}^{2} p_k \varphi(x; m_k, \sigma_k)$$

with

$$\varphi(x; m_k, \sigma_k) = \frac{1}{\sqrt{(2\pi)^D} \sigma_k} e^{-\frac{1}{2}\left(\frac{\|x - m_k\|}{\sigma_k}\right)^2}$$

where p_k are the mixing probabilities and m_k and σ_k are the mean and variance of the kth distribution, respectively. The number of dimensions is denoted by D ($D = 2$). The method yielded the parameters $\vartheta = (p_k, \mu_k, \sigma_k)$ that are most likely to have generated the data. We used a slightly higher threshold (discrimination: $CS^+/CS^- > 1.4$) than the one estimated by the expectation-maximization algorithm, to decrease the false positives and thus to have a more conservative estimate of the number of discriminating mice (Supplementary Fig. 16).

To test whether changes in tonic baseline firing rates of CEl neurons can predict generalization versus discrimination, we performed a receiver operating characteristic (ROC) analysis using the classification results of the behavioural data. The area under the ROC curve (AUC) had a value of 0.67, which is above chance level (0.5). To test significance of this value we computed AUC values for 1,000 surrogate samples, which were random permutations of the initial classification scheme. The results revealed significance ($P < 0.01$). Note, that significant results were also obtained using an equivalent Wilcoxon signed-rank test ($P < 0.01$). Thus, changes in tonic activity are a useful discriminator for freezing behaviour revealing that CEl neurons in generalizing mice exhibit higher changes (less negative and more positive) in tonic activity than CEl neurons in discriminating mice.

Surgery and single-unit recordings. Mice were anaesthetized with isoflurane (induction 5%, maintenance 2.5%) in O_2. Body temperature was maintained with a heating pad (CMA/150, CMA/Microdialysis). Mice were secured in a stereotaxic frame and unilaterally implanted in the amygdala with a multi-wire electrode aimed at the following coordinates: 1.3 mm posterior to bregma; ±2.9 mm lateral to midline; and 3.9 mm to 4.3 mm deep from the cortical surface. The electrodes consisted of 8–16 individually insulated nichrome wires (13 μm inner diameter, impedance 50–300 kΩ; California Fine Wire) contained in a 26-gauge stainless steel guide canula. The wires were attached to a 10 pin to 18 pin connector (Omnetics). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days. Analgesia was applied before and during the 3 days after surgery (Metacam). Electrodes were connected to a headstage (Plexon) containing eight to sixteen unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain ×100, band-pass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz and band-pass filtered from 250 Hz to 8 kHz, and was isolated by time–amplitude window discrimination and template matching using a Multichannel Acquisition Processor system (Plexon). At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques.

Single-unit spike sorting and analysis. Single-unit spike sorting was performed using an Off-Line Spike Sorter (Plexon) as described[21,44,45] (Supplementary Fig. 14). Principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to be generated from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units and if it displayed a clear refractory period (>1 ms) in the auto-correlogram histograms. In addition, two parameters were used to quantify the overall separation between identified clusters in a particular channel. These parameters include the J3 statistic, which corresponds to the ratio of between-cluster to within-cluster scatter, and the Davies–Bouldin validity index, which reflects the ratio of the sum of within-cluster scatter to between-cluster separation[44]. High values for the J3 and low values for the Davies–Bouldin validity index are indicative of good cluster separation. Control values for these statistics were obtained by artificially defining two clusters from the centred cloud of points in the principal component space from channels in which no units could be detected. Template waveforms were then calculated for well-separated clusters and stored for further analysis. Clusters of identified neurons were analysed offline for each recording session using principal component analysis and a template-matching algorithm. Only stable clusters of single units recorded over the time course of the entire behavioural training were considered.

To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fires, only one of the two neurons was considered for further analysis. CS-induced neural activity was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size, 50 ms; averaged over blocks of four CS presentations consisting of 108 individual sound pips in total) using a z-score transformation. Z-score values were calculated by subtracting the average baseline firing rate established over the 500-ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Classification of units was performed by considering a significant z-score value within 200 ms after CS onset during fear test. For statistical analysis, z-score comparisons were performed using the average z-score value calculated during the 100 ms after CS onset. Tonic activity at test was z-scored by calculating the average firing rate ±s.d. of 108 randomly chosen 500-ms sweeps during the pre-CS period and z-scoring tonic activity during the 500-ms pre-pip period to the pre-CS period. Results are presented as mean ± s.e.m.

To address CS-evoked latencies of the three CEA neuronal populations, normalized peri-stimulus time histograms (PSTH) were computed for each single neuron of each category using 5-ms bins. Population PSTHs were obtained by averaging single neuron PSTHs. CS-evoked onset latencies were calculated for the population PSTH based on the first significant bin (at least 2.5 s.d. of baseline activity).

To assess the significance of cross-correlograms during spontaneous activity between a reference and a target neuron, mean firing rate with 95% confidence limits of the target neuron was calculated. Short-latency inhibitory cross-correlograms were considered to be significant if the number of action potentials of the target neuron (-50 ms to 50 ms) was inferior to the 95% confidence limits. Furthermore, to show that the cross-correlograms were not simply occurring

by chance, the spike train of the target neuron was shuffled 100 times and a shuffled cross-correlogram was computed[46]. Absence of short-latency interaction in the shuffled cross-correlogram was indicative that the cross-correlations were not due to chance.

Muscimol iontophoresis. Muscimol microiontophoresis was performed in chronically implanted animals[21]. Single-barrel micropipettes with a tip diameter of 10 to 15 μm were cut at 1-cm length and filled with a solution containing muscimol covalently coupled to a fluorophore (Muscimol-Bodipy-TMR conjugated, Invitrogen; 5 mM in phosphate buffered saline (PBS) 0.1 M, DMSO 40%) or with bodipy alone (Invitrogen; 5 mM in PBS 0.1 M, DMSO 40%). Mice were bilaterally implanted at the following coordinates: 1.3 mm posterior to bregma; 2.9 mm lateral to midline; and 3.9 mm to 4.3 mm deep from the cortical surface. Chlorided silver wires were inserted in each micropipette and attached to a connector. A third silver wire screwed onto the skull and attached to the connector served as a reference electrode. The entire miniature was secured using cyanoacrylate adhesive gel. After surgery, mice were allowed to recover for 2 days. On the injection day, iontophoretic applications were performed by means of cationic current ($+12\,\mu A$ to $+15\,\mu A$) for 15 min per side using a precision current source device (Stoelting). Mice were submitted to the behavioural procedure 5 min after the end of iontophoretic injections and were immediately perfused at the end of the experiments. Brains were collected for further histological analysis. Serial slices containing the amygdala were imaged at ×5 using an epifluorescence stereo microscope (Leica Microsystems), and the location and the extent of the injections were controlled. Statistical analyses were performed using paired and unpaired Student's *t*-test post-hoc comparisons at the $P < 0.05$ level of significance. Results are presented as mean ± s.e.m.

Intracellular recordings and morphological reconstructions. Intracellular recording sessions were done in mice under chloral hydrate anaesthesia (400 mg kg^{-1}), and ended the same day with the animal being transcardially perfused and the brain kept for morphological reconstruction of the neurobiotin-filled recorded neurons using standard methods[47]. During the experiment, the animal's head was held firmly by a holding bar cemented on the cranium. The absence of ear-bars allowed the use of an open-field speaker (ES1 Free Field Electrostatic Speaker, TDT) for auditory stimulation. Auditory responses of CEl neurons were determined by the presentation of tones of different frequencies (1–30 kHz) and intensities (using a RP2.1 processor and a HB7 headphone driver; TDT).

Intracellular electrodes were pulled from borosilicate glass tubing (1.5 mm outer diameter, 0.84 mm inner diameter; World Precision Instruments) using a Flaming-Brown micropipette puller (model P-97; Sutter Instruments). Electrodes were filled with 1.5% neurobiotin (Vector Laboratories Inc.) in 1 M potassium acetate. Impedances were measured *in situ* and ranged from 65 to 120 MΩ. Electrodes were slowly lowered in the brain via a micromanipulator (LN mini/combi; Luigs & Neumann). Recordings were acquired and analysed with ClampEx9.0 and ClampFit9.0 (Molecular Devices) through an intracellular recording amplifier (Axoclamp-2B, Molecular Devices) and a data digitizer (Digidata 1322A). Positive DC pulses (0.1–1.0 nA, 500 ms, 1 Hz) were used to eject neurobiotin into the neurons. Mice were then perfused transcardially with 4% paraformaldehyde in 0.1 M phosphate buffer. Brains were removed and stored in the perfusion fixative. They were later sliced on a microtome into 80-μm-thick sections and labelled for neurobiotin using the Vectastain Elite avidin–biotin complex peroxidase kit (Vector Laboratories Inc.). Neurons were reconstructed with the Neurolucida software (Microbrightfield).

Virus injections and optical stimulation. For optical activation of CEm output neurons, animals were injected into CEm with an AAV serotype 2/7 (Vector Core), containing a construct coding for ChR22A-tdimer[20] at -1.4 mm posterior and ± 2.9 mm lateral to bregma at a depth of -4.4 mm. Briefly, deeply anaesthetized animals were fixed in a stereotactic frame (Kopf Instruments) and the skin above the skull was cut. Glass pipettes (tip diameter 10–20 μm), connected to a Picospritzer III (Parker Hannifin Corporation), were lowered by a Micropositioner (Kopf Instruments) to the depth of 4.4 mm. About 300 nl were pressure injected into CEm. In the same surgeries 26-gauge stainless steel guide cannulae (Plastics One) were implanted bilaterally along the same track above CEm at a depth of -3.9 mm. Guide cannulae were secured using cyanoacrylate adhesive gel (Henkel) and dental cement (Heraeus Dental). To prevent blockage of the cannulae, dummy cannulae (Plastics One) were inserted and fixed. Behavioural experiments were performed after 4 weeks of recovery and expression time and 3 days of handling. Dummy cannulae were removed and optic fibres with a diameter of 200 μm (Thorlabs GmbH) were inserted bilaterally into the implanted guide cannulae. Mice were then placed into a behavioural context and the optic fibres were connected to a blue laser ($\lambda = 473$ nm, 100 mW, Extreme Lasers). The mice received four 10-s pulses of blue light with intervals between 20 s and 60 s. Freezing with and without light stimulation was quantified as described. After the experiment, optic fibres were removed and animals were perfused for histological analysis of the injection site as described.

For retrograde tracing of projections, replication defective herpes simplex virus (HSV-1) (BioVex)[48] expressing eGFP was injected into either CEl (anteroposteriorly, -1.2 mm; laterality, -2.9 mm; depth, 4.1 mm) or CEm (see above). For identification of the injection site, the virus solution was mixed at 1:1,000 with blue fluorescing polymer microspheres (Duke Scientific Corp.). Before and after the surgery, systemic (Metacam, Boehringer Ingelheim) and local analgesic (Naropin, AstraZeneca AG) were administered. After 1 week of expression, animals were transcardially perfused with 4% PFA. The brain was removed and cut into 80 μm coronal slices. To improve the fluorescent signal, an immunostaining was performed. Slices were kept in blocking solution (3% BSA, 0.2% Triton in 0.1M PBS) for 1 h at room temperature, before application of the primary antibody (goat anti-GFP, Abcam; 1:500 in blocking solution) and incubated at 4 °C over night. After washing, slices were incubated with secondary antibody (Alexa Fluor 488, donkey anti goat, Invitrogen; 1:1,000 in PBS) at 4 °C over night. After a final wash, slices were mounted on cover slips and imaged.

45. Herry, C. *et al.* Processing of temporal unpredictability in human and animal amygdala. *J. Neurosci.* **27,** 5958–5966 (2007).
46. Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsaki, G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neurosci.* **11,** 823–833 (2008).
47. Lang, E. J. & Paré, D. Synaptic and synaptically activated intrinsic conductances underlie inhibitory potentials in cat lateral amygdaloid projection neurons *in vivo. J. Neurophysiol.* **77,** 353–363 (1997).
48. Lima, S. Q., Hromadka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS ONE* **4,** e6099 (2009).

LETTER

Hidden magnetic excitation in the pseudogap phase of a high-T_c superconductor

Yuan Li[1]†, V. Balédent[2], G. Yu[3], N. Barišić[4,5]†, K. Hradil[6], R. A. Mole[7], Y. Sidis[2], P. Steffens[8], X. Zhao[4,9], P. Bourges[2] & M. Greven[3]

The elucidation of the pseudogap phenomenon of the high-transition-temperature (high-T_c) copper oxides—a set of anomalous physical properties below the characteristic temperature T^* and above T_c—has been a major challenge in condensed matter physics for the past two decades[1]. Following initial indications of broken time-reversal symmetry in photoemission experiments[2], recent polarized neutron diffraction work demonstrated the universal existence of an unusual magnetic order below T^* (refs 3, 4). These findings have the profound implication that the pseudogap regime constitutes a genuine new phase of matter rather than a mere crossover phenomenon. They are furthermore consistent with a particular type of order involving circulating orbital currents, and with the notion that the phase diagram is controlled by a quantum critical point[5]. Here we report inelastic neutron scattering results for $HgBa_2CuO_{4+\delta}$ that reveal a fundamental collective magnetic mode associated with the unusual order, and which further support this picture. The mode's intensity rises below the same temperature T^* and its dispersion is weak, as expected for an Ising-like order parameter[6]. Its energy of 52–56 meV renders it a new candidate for the hitherto unexplained ubiquitous electron–boson coupling features observed in spectroscopic studies[7–10].

Our measurements were performed on three samples made of co-aligned crystals, which were grown by a self-flux method[11] and were free from substantial macroscopic impurity phases and inhomogeneity (Supplementary Information sections 1 and 2). $HgBa_2CuO_{4+\delta}$ (Hg1201) exhibits the highest value of T_c of all copper oxides with one copper–oxygen plane per unit cell, has a simple tetragonal structure, and is furthermore thought to be relatively free of disorder effects[12,13]. Here we quote the scattering wave vector as $\mathbf{Q} = H\mathbf{a^*} + K\mathbf{b^*} + L\mathbf{c^*} \equiv (H,K,L)$ in reciprocal lattice units (r.l.u.), where $\mathbf{a^*}$, $\mathbf{b^*}$ and $\mathbf{c^*}$ are reciprocal lattice vectors. We present normalized neutron intensities in most figures to facilitate a direct comparison of the intensity among the measurements (the normalization process is described in Supplementary Information section 3).

Spin-polarized inelastic neutron scattering data (Fig. 1) demonstrate the existence of a magnetic excitation throughout the two-dimensional (2D) Brillouin zone in a nearly optimally doped sample ($T_c = 94.5 \pm 2$ K, denoted OP95). Energy scans in the spin-flip channel reveal a resolution-limited feature at low temperatures, with a weak dispersion and a maximum of 56 meV at the 2D zone-corner ($H = K = 0.5$, also referred to as the antiferromagnetic wave vector, $\mathbf{q}_{AF}$). The feature cannot be due to a polarization leakage from the non-spin-flip channel (Supplementary Information section 4), and it disappears in the spin-flip channel at 300 K (Fig. 1a). Background intensity at 10 K has been measured separately using a combination of different spin-polarization geometries (Supplementary Information section 1) and agrees with the intensity at 300 K within the error (Fig. 1a; $H = K = 0.2$). These results prove the magnetic origin of the peak at 10 K.

The dispersion of the excitation along (H, H) for both this optimally doped sample and an underdoped sample ($T_c = 65 \pm 3$ K, UD65; Supplementary Information section 5), measured with both polarized and unpolarized neutrons, is displayed in Fig. 1c. The weak dispersion
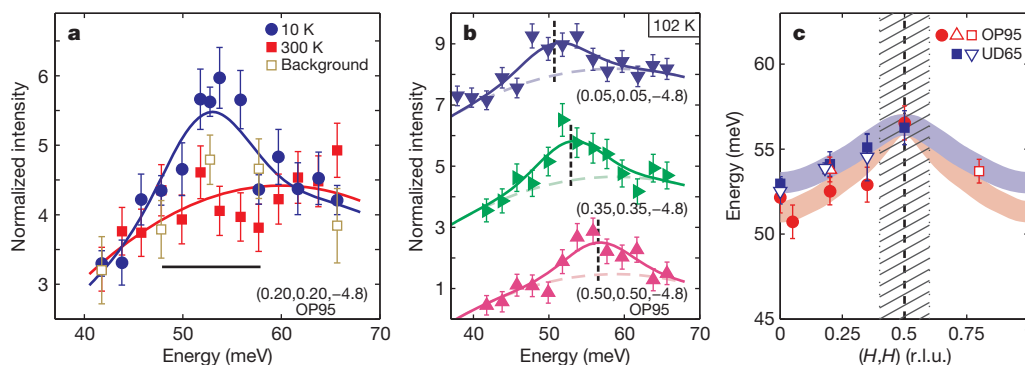
Figure 1 | Identification of a weakly dispersing magnetic collective mode. **a**, Spin-flip energy scans for sample OP95; **Q**-position given at bottom right. Background (open squares) is measured at 10 K by a method described in Supplementary Information section 1 and approximated together with the data at 300 K by a parabolic baseline (red line). The 10 K data are fitted to a Gaussian (blue line) on this baseline, with a small offset to account for the possible background change with temperature. Similar baselines are used in **b** and Fig. 2a. Horizontal bar indicates instrument energy resolution of ~10 meV (FWHM). **b**, Spin-flip energy scans at additional **Q**-positions, offset for clarity.

c, Summary of dispersion along (H,H). Different symbols represent measurements using different spectrometers: IN20 (circles), PUMA (squares), 2T (triangle) and IN8 (reversed triangle). The measurement on spectrometer IN20 is spin-polarized; all others are unpolarized. Data are presented in panels **a** and **b**, Fig. 2c and d, and Supplementary Figs 4–6. Conventional magnetic response near $\mathbf{q}_{AF}$ (vertical dashed line) is expected in the hatched area (estimated on the basis of Supplementary Fig. 8b–d), where the determination of the dispersion using energy scans may be less accurate. Error bars represent statistical and fit uncertainties (1 s.d.).

[1]Department of Physics, Stanford University, Stanford, California 94305, USA. [2]Laboratoire Léon Brillouin, CEA-CNRS, CEA-Saclay, 91191 Gif sur Yvette, France. [3]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA. [4]T.H. Geballe Laboratory for Advanced Materials, Stanford University, Stanford, California 94305, USA. [5]1. Physikalisches Institut, Universität Stuttgart, 70550 Stuttgart, Germany. [6]Institut für Physikalische Chemie, Universität Göttingen, 37077 Göttingen, Germany. [7]Forschungsneutronenquelle Heinz Maier-Leibnitz, 85747 Garching, Germany. [8]Institut Laue Langevin, 38042 Grenoble Cedex 9, France. [9]State Key Laboratory of Inorganic Synthesis and Preparative Chemistry, College of Chemistry, Jilin University, Changchun 130012, China. †Present addresses: Max Planck Institute for Solid State Research, 70569 Stuttgart, Germany (Y.L.); Institute of Physics, Bijenicka cesta 46, 10 000 Zagreb, Croatia (N.B.).

(<10%) and the strong response at the 2D zone centre $q = 0$ differ drastically from the characteristics of the well-known antiferromagnetic response near $\mathbf{q}_{AF}$ (refs 14, 15). Remarkably, the dispersion of the excitation, which is already present well above T_c, reaches its maximum at the same point in energy–momentum space as the so-called magnetic resonance[16], which in OP95 occurs only below T_c (ref. 17). This is further demonstrated in Fig. 2a: compared to the measurement above T_c, substantially higher intensity is observed at $\mathbf{q}_{AF}$ below T_c.

We emphasize that the magnetic signal far away from $\mathbf{q}_{AF}$ cannot be attributed to a resonance peak that is broad in momentum, for the following reasons. First, at optimal doping, the temperature dependence of the signal away from $\mathbf{q}_{AF}$ (Fig. 3a) is different from that of the resonance[17] (Fig. 2a inset). Second, the excitation energy near $q = 0$ differs from that at $\mathbf{q}_{AF}$ (Fig. 1, Supplementary Fig. 4). Third, the profile of momentum scans at the resonance energy is not symmetric about $\mathbf{q}_{AF}$, but is better described by a broad peak centred at $q = 0$ plus a narrower peak centred at $\mathbf{q}_{AF}$ (Supplementary Fig. 8a). Fourth, the resonance peak in momentum scans does not extend below $H = K = 0.3$ (Supplementary Fig. 8b–d). Therefore, a magnetic excitation branch in addition to the resonance is required to describe the data, as illustrated in Fig. 1c. This excitation branch is also distinctly different from the well-known 'hourglass' excitations[18,19]: the latter only exist in a limited momentum range near $\mathbf{q}_{AF}$ (the hatched area in Fig. 1c) and only become clearly incommensurate[15] below T_c in $YBa_2Cu_3O_{6+\delta}$ (YBCO), whereas the former is observed all the way to $q = 0$ and above T_c (Fig. 1b and Fig. 3). Moreover, following the notion that the hourglass excitations are collective modes below the electron–hole continuum, they are expected only near $\mathbf{q}_{AF}$ and cannot continuously disperse to $q = 0$ (ref. 19).

After the magnetic nature of the excitation was verified with polarized neutrons, further quantitative measurements were carried out with unpolarized neutrons to benefit from the much higher neutron flux.

Following standard procedure to extract a magnetic signal[19], phonons and spurious contributions were either removed by subtracting background obtained at high temperature, or avoided by carefully choosing the measurement conditions (Supplementary Information sections 5 and 6). Measurements at 2D $\mathbf{Q}$-positions similar to those in Fig. 1a, b, shown in Supplementary Figs 4–6, confirm and extend the spin-polarized results. The excitation was also observed at $\mathbf{q} = (0.5,0)$ (Fig. 2b) and $\mathbf{q} = (0,0.5)$ (Fig. 2c), which are equivalent 2D $\mathbf{Q}$-positions rotated 45° away from those summarized in Fig. 1c. The energy width of the excitation was found to remain resolution-limited when measured with better energy resolution (Fig. 2d), which indicates that it is a long-lived mode. As the excitation is observed at all of those $\mathbf{Q}$-positions summarized in Fig. 2d inset, we conclude that it is present throughout the entire 2D Brillouin zone.

The temperature dependence of the excitation is best measured away from $\mathbf{q}_{AF}$ and with unpolarized neutrons (Supplementary Information section 6). The results are summarized in Fig. 3a. The onset temperature of the excitation, T_{ex}, is shown in Fig. 3b together with T^* determined from in-plane resistivity[12,20] (Supplementary Information section 7) and the onset temperature of the $q = 0$ magnetic order measured by polarized-neutron diffraction[4]. The good agreement among these results suggests that the excitation is a fundamental collective mode of the universal $q = 0$ pseudogap order[3,4]. As the high-T_c copper oxides are not ferromagnetic, a 'decoration' of the unit cell with a net cancellation of moments is required to account for the observed behaviour. Consequently, the collective mode can not be understood with conventional t-J and one-band Hubbard models[21], which reduce the problem to one site per unit cell, and instead an extended multi-band approach appears necessary[5,22,23].

The unusual phase diagram of the copper oxides has been argued to be controlled by an underlying quantum critical point that marks the termination of a distinct order[5,24–26]. A leading candidate is the $q = 0$

Figure 2 | Presence of the collective mode throughout the entire 2D Brillouin zone. a, Spin-flip energy scans at $\mathbf{q}_{AF}$ below and above T_c. Arrow and dashed line indicate the estimated intensity change due to the resonance. The inset is adapted from ref. 17 and illustrates the intensity change of the resonance below T_c. b, Unpolarized spectral difference between 16 K and 200 K for sample UD89 ($T_c = 89 \pm 3$ K) at a $\mathbf{Q}$-position away from the 2D zone diagonal. c, Unpolarized spectral difference between 4 K and 330 K for sample UD65 at three different $\mathbf{Q}$-positions in the first 2D Brillouin zone. d, Unpolarized measurement for sample UD65 using better energy resolution. A constant has

been subtracted from the 403 K data for better comparison. The peak at 5 K is no longer present at 403 K, which is well above T^*. Lines in all panels are Gaussian fits, which serve as guides to the eye. Open symbols in d indicate measurement points that seem contaminated by phonons (below 50 meV) and a spurious contribution (at 53 meV). Inset in d summarizes the $\mathbf{Q}$-positions (colour-coded for the main panels) at which the measurements were performed. Horizontal bars indicate energy resolutions of the instruments (FWHM) and error bars represent statistical uncertainty (1 s.d.).

Figure 3 | Temperature dependence of the collective mode demonstrates its connection to the pseudogap phenomenon. a, Temperature dependence of intensity measured at 53 meV, $\mathbf{Q} = (0.2,0.2,5.2)$, for sample OP95 (triangles) and at 54 meV, $\mathbf{Q} = (0.2,0.2,6.0)$, for sample UD65 (circles), after background subtraction (Supplementary Information section 6) and normalization to values at the lowest temperature. Lines are empirical power-law fits (Supplementary Information section 6). The onset temperature T_{ex} is 211 ± 13 K near optimal doping and becomes considerably higher (335 ± 23 K) at the lower doping, with no abrupt change near T_c in either case. **b**, Summary of characteristic (onset) temperatures. Red circles, excitation branch (this work); blue squares, $q = 0$ magnetic order[4]; green triangles, in-plane resistivity deviation (refs 12, 20 and Supplementary Information section 7). Hole concentrations are determined after ref. 29 based on the doping dependence of T_c in Hg1201 (black line). Error bars, 1 s.d.

magnetic order that preserves the translational symmetry of the lattice[3,4] and which would naturally give rise to excitations centred at $q = 0$. Indeed, such an ordered state involving circulating charge currents has been predicted theoretically[5,27]. On the basis that this current-loop order is describable by an Ising-like Ashkin-Teller model, a rather unusual magnetic excitation spectrum with nearly dispersion-free excitations is expected from the discrete symmetry of the order parameter[6], consistent with our findings.

In the copper oxides, anomalies in the charge excitation spectrum are usually discussed in terms of a coupling between electrons and bosonic modes (phonons or antiferromagnetic spin fluctuations). The hitherto unobserved excitation found here at the same energy as the resonance, but up to higher temperature and all the way to $q = 0$, is a new candidate for the mysterious electron–boson interaction features observed by photoemission[7], optical spectroscopy[8,9] and scanning tunnelling spectroscopy[10]. At $\mathbf{q}_{AF}$, the strength of the excitation is comparable to that of the resonance (in Hg1201 (Fig. 2a) and YBCO[19]). Whereas the latter is located at $\mathbf{q}_{AF}$, the former extends throughout the entire Brillouin zone (Fig. 2). As the area of resolution ellipsoid in 2D momentum space is a few per cent of the Brillouin zone, we estimate that the momentum-integrated spectral weight of the excitation branch is at least an order of magnitude greater than that of the resonance in Hg1201, and comparable to that of the full antiferromagnetic response in underdoped YBCO (the integrated spectral weight between 25 and 100 meV is believed to be several times larger than that of the resonance[15,28]). In other words, about half of the total magnetic spectral weight is located within a narrow range around the resonance energy, and has been hidden so far, in part due to the excitation's weak momentum dependence. It remains an open question whether the coincidence of energy scales of the excitation and resonance is accidental, or if there is a profound physical connection. Recently, we observed indications of the existence of a second branch of similar excitations at lower energy, which needs to be verified by further studies.

All the evidence[12,13] suggests that Hg1201 is not only representative of the copper oxides, but is a model compound, and therefore experiments on Hg1201 can be expected to reveal the essence of the underlying physics most clearly. Given the universal existence among the copper oxides of the pseudogap phase, of the $q = 0$ magnetic order, and of the

electron–boson coupling features in the 50–60 meV range, we expect the excitation branch to be present in other copper oxides as well (Supplementary Information section 9).

1. Norman, M. R., Pines, D. & Kallin, C. The pseudogap: friend or foe of high T_c? *Adv. Phys.* **54**, 715–733 (2005).
2. Kaminski, A. *et al.* Spontaneous breaking of time-reversal symmetry in the pseudogap state of a high-T_c superconductor. *Nature* **416**, 610–613 (2002).
3. Fauqué, B. *et al.* Magnetic order in the pseudogap phase of high-T_c superconductors. *Phys. Rev. Lett.* **96**, 197001 (2006).
4. Li, Y. *et al.* Unusual magnetic order in the pseudogap region of the superconductor HgBa$_2$CuO$_{4+\delta}$. *Nature* **455**, 372–375 (2008).
5. Varma, C. M. Non-Fermi-liquid states and pairing instability of a general model of copper oxide metals. *Phys. Rev. B* **55**, 14554–14580 (1997).
6. Varma, C. M. Theory of the pseudogap state of the cuprates. *Phys. Rev. B* **73**, 155113 (2006).
7. Lanzara, A. *et al.* Evidence for ubiquitous strong electron-phonon coupling in high-temperature superconductors. *Nature* **412**, 510–514 (2001).
8. Yang, J. *et al.* Exchange boson dynamics in cuprates: optical conductivity of HgBa$_2$CuO$_{4+\delta}$. *Phys. Rev. Lett.* **102**, 027003 (2009).
9. van Heumen, E. *et al.* Optical determination of the relation between the electron-boson coupling function and the critical temperature in high-T_c cuprates. *Phys. Rev. B* **79**, 184512 (2009).
10. Lee, J. *et al.* Interplay of electron-lattice interactions and superconductivity in Bi$_2$Sr$_2$CaCu$_2$O$_{8+\delta}$. *Nature* **442**, 546–550 (2006).
11. Zhao, X. *et al.* Crystal growth and characterization of the model high-temperature superconductor HgBa$_2$CuO$_{4+\delta}$. *Adv. Mater.* **18**, 3243–3247 (2006).
12. Barišić, N. *et al.* Demonstrating the model nature of the high-temperature superconductor HgBa$_2$CuO$_{4+\delta}$. *Phys. Rev. B* **78**, 054518 (2008).
13. Eisaki, H. *et al.* Effect of chemical inhomogeneity in bismuth-based copper oxide superconductors. *Phys. Rev. B* **69**, 064512 (2004).
14. Vignolle, B. *et al.* Two energy scales in the spin excitations of the high-temperature superconductor La$_{2-x}$Sr$_x$CuO$_4$. *Nature Phys.* **3**, 163–167 (2007).
15. Hinkov, V. *et al.* Spin dynamics in the pseudogap state of a high-temperature superconductor. *Nature Phys.* **3**, 780–785 (2007).
16. Rossat-Mignod, J. *et al.* Neutron scattering study of the YBa$_2$Cu$_3$O$_{6+x}$ system. *Physica C* **185–189**, 86–92 (1991).
17. Yu, G. *et al.* Magnetic resonance in the model high-temperature superconductor HgBa$_2$CuO$_{4+\delta}$. *Phys. Rev. B* **81**, 064518 (2010).
18. Tranquada, J. M. *et al.* Quantum magnetic excitations from stripes in copper oxide superconductors. *Nature* **429**, 534–538 (2004).
19. Pailhès, S. *et al.* Resonant magnetic excitations at high energy in superconducting YBa$_2$Cu$_3$O$_{6.85}$. *Phys. Rev. Lett.* **93**, 167001 (2004).
20. Grbić, M. S. *et al.* Microwave measurements of the in-plane and c-axis conductivity in HgBa$_2$CuO$_{4+\delta}$: discriminating between superconducting fluctuations and pseudogap effects. *Phys. Rev. B* **80**, 094511 (2009).
21. Norman, M. R. & Pépin, C. The electronic nature of high temperature cuprate superconductors. *Rep. Prog. Phys.* **66**, 1547–1610 (2003).
22. Tahir-Kheli, J. & Goddard, W. A. III. Chiral plaquette polaron theory of cuprate superconductivity. *Phys. Rev. B* **76**, 014514 (2007).
23. Weber, C. *et al.* Orbital currents in extended Hubbard models of high-T_c cuprate superconductors. *Phys. Rev. Lett.* **102**, 017005 (2009).
24. Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
25. Sachdev, S. Quantum criticality: competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
26. Chakravarty, S. *et al.* Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503 (2001).
27. Simon, M. E. & Varma, C. M. Detection and implications of a time-reversal breaking state in underdoped cuprates. *Phys. Rev. Lett.* **89**, 247003 (2002).
28. Hayden, S. M. *et al.* The structure of the high-energy spin excitations in a high-transition-temperature superconductor. *Nature* **429**, 531–534 (2004).
29. Yamamoto, A., Hu, W. & Tajima, S. Thermoelectric power and resistivity of HgBa$_2$CuO$_{4+\delta}$ over a wide doping range. *Phys. Rev. B* **63**, 024504 (2000).

LETTER

Ultrathin compound semiconductor on insulator layers for high–performance nanoscale transistors

Hyunhyub Ko[1,2,3]*†, Kuniharu Takei[1,2,3]*, Rehan Kapadia[1,2,3]*, Steven Chuang[1,2,3], Hui Fang[1,2,3], Paul W. Leu[1,2,3], Kartik Ganapathi[1], Elena Plis[5], Ha Sul Kim[5], Szu–Ying Chen[4], Morten Madsen[1,2,3], Alexandra C. Ford[1,2,3], Yu–Lun Chueh[4], Sanjay Krishna[5], Sayeef Salahuddin[1] & Ali Javey[1,2,3]

Over the past several years, the inherent scaling limitations of silicon (Si) electron devices have fuelled the exploration of alternative semiconductors, with high carrier mobility, to further enhance device performance[1–8]. In particular, compound semiconductors heterogeneously integrated on Si substrates have been actively studied[7,9,10]: such devices combine the high mobility of III–V semiconductors and the well established, low-cost processing of Si technology. This integration, however, presents significant challenges. Conventionally, heteroepitaxial growth of complex multilayers on Si has been explored[9,11–13]—but besides complexity, high defect densities and junction leakage currents present limitations in this approach. Motivated by this challenge, here we use an epitaxial transfer method for the integration of ultrathin layers of single-crystal InAs on Si/SiO$_2$ substrates. As a parallel with silicon-on-insulator (SOI) technology[14], we use 'XOI' to represent our compound semiconductor-on-insulator platform. Through experiments and simulation, the electrical properties of InAs XOI transistors are explored, elucidating the critical role of quantum confinement in the transport properties of ultrathin XOI layers. Importantly, a high-quality InAs/dielectric interface is obtained by the use of a novel thermally grown interfacial InAsO$_x$ layer (~1 nm thick). The fabricated field-effect transistors exhibit a peak transconductance of ~1.6 mS μm^{-1} at a drain–source voltage of 0.5 V, with an on/off current ratio of greater than 10,000.

Epitaxial lift-off and transfer of crystalline microstructures to various support substrates has been shown to be a versatile technique for applications ranging from optoelectronics to large-area electronics[15–18]. Specifically, high-performance, mechanically flexible macro-electronics and photovoltaics have been demonstrated on plastic, rubber and glass substrates by this method[19–21]. Here we use a modified epitaxial transfer scheme for integrating ultrathin InAs layers (with nanometre-scale thicknesses) on Si/SiO$_2$ substrates for use as high-performance nanoscale transistors. These InAs layers are fully depleted, which is an important criterion for achieving high-performance field-effect transistors (FETs) with respectable 'off' currents based on small bandgap semiconductors. The transfer is achieved without the use of adhesive layers, thereby allowing the use of purely inorganic interfaces with low interface trap densities and high stability. Figure 1a shows a diagram of the fabrication process for InAs XOI substrates (see Methods for details).

We used atomic force microscopy (AFM) to characterize the surface morphology and uniformity of the fabricated XOI substrates. Figure 1b and c shows representative AFM images of an array of InAs nanoribbons (~18 nm thick) on a Si/SiO$_2$ substrate, clearly depicting the smooth surfaces (<1 nm surface roughness) and high uniformity of the enabled structures over large areas. Uniquely, the process readily enables the heterogeneous integration of different III–V materials and structures on a single substrate through a multi-step epitaxial transfer

process. To demonstrate this capability, a two-step transfer process was used to form ordered arrays of 18- and 48-nm-thick InAs nanoribbons that are perpendicularly oriented on the surface of a Si/SiO$_2$ substrate (Fig. 1d, e). This result demonstrates the potential capacity of the proposed XOI technology for generic heterogeneous and/or hierarchical assembly of crystalline semiconducting materials. In the future, a similar scheme may be used to enable the fabrication of both p- and n- type transistors on the same chip for complementary electronics based on the optimal III–V semiconductors.
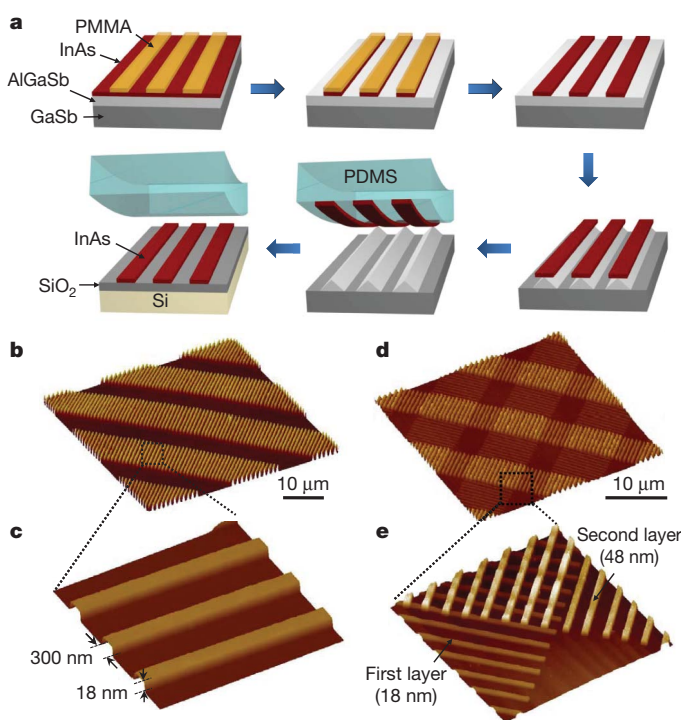
Figure 1 | Fabrication scheme for ultrathin InAs XOI, and AFM images.
a, Schematic procedure for the assembly of InAs XOI substrates by an epitaxial transfer process. The epitaxially grown, single-crystal InAs films are patterned with PMMA and wet etched into nanoribbon arrays. A subsequent selective wet etch of the underlying AlGaSb layer and the transfer of nanoribbons by using an elastomeric PDMS slab result in the formation of InAs nanoribbon arrays on Si/SiO$_2$ substrates. **b, c,** AFM images of InAs nanoribbon arrays on a Si/SiO$_2$ substrate. The nanoribbons are ~10 μm long, 18 nm high and ~300 nm wide. **d, e,** AFM images of InAs nanoribbon superstructures on a Si/SiO$_2$ substrate, consisting of two layers of perpendicularly oriented nanoribbon arrays with 18- and 48-nm thicknesses, as assembled by a two-step epitaxial transfer process.

[1]Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [3]Berkeley Sensor and Actuator Center, University of California, Berkeley, California 94720, USA. [4]Materials Science and Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan. [5]Electrical and Computer Engineering Department, and Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106, USA. †Present address: School of Nanotechnology and Chemical Engineering, Ulsan National Institute of Science and Technology, Ulsan Metropolitan City, South Korea.
*These authors contributed equally to this work.

To shed light on the atomic structure of the interfaces, cross-sectional transmission electron microscopy (TEM) images of an InAs XOI device were taken and are shown in Fig. 2. The high-resolution TEM (HRTEM) image (Fig. 2c) illustrates the single-crystal structure of the InAs nanoribbons (~13 nm thick) with atomically abrupt interfaces with the SiO$_2$ and ZrO$_2$ layers. The TEM image of the InAs/SiO$_2$ interface does not exhibit visible voids (Fig. 2c), although only a small fraction of the interface is examined by TEM. As described in more depth below, InAs nanoribbons were thermally oxidized before the top-gate stack deposition to drastically lower the interfacial trap densities. The thermally grown InAsO$_x$ layer is clearly evident in the HRTEM image (Fig. 2c), with a thickness of ~1 nm.

Long-channel, back-gated FETs based on individual nanoribbons were fabricated in order to elucidate the intrinsic electron transport properties of InAs nanoribbons as a function of thickness. The process scheme involved the fabrication of XOI substrates with the desired InAs thickness, followed by the formation of source/drain (S/D) metal contacts by lithography and lift-off (~50-nm-thick Ni). The p$^+$ Si support substrate was used as the global back-gate, with a 50-nm thermal SiO$_2$ layer as the gate dielectric. Nickel contacts were annealed at 225 °C for 5 min in N$_2$ to enable the formation of low-resistance contacts to the conduction band of InAs (Supplementary Fig. 6)[22]. The transfer characteristics (at a drain–source voltage (V_{DS}) of 0.1 V) of the back-gated XOI FETs with a channel length $L \approx 5\,\mu$m and InAs thicknesses of 8–48 nm are shown in Fig. 3a. Two trends are clearly evident
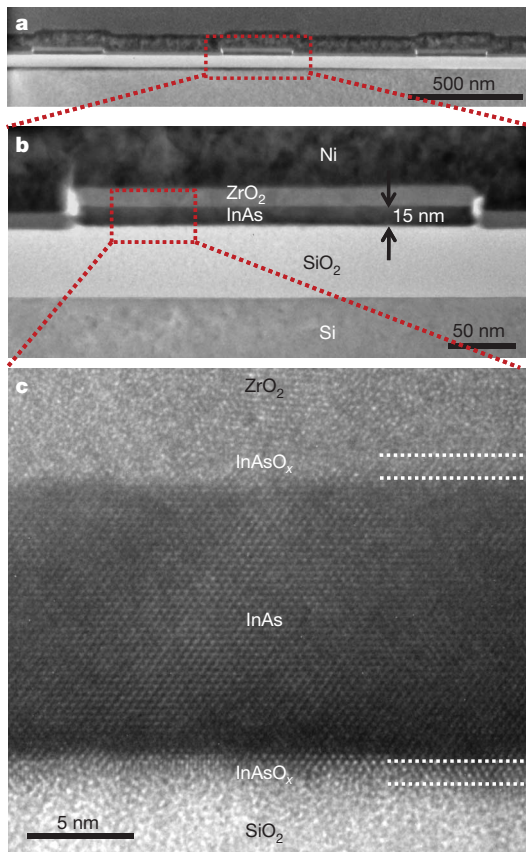
Figure 2 | Cross-sectional TEM analysis of InAs XOI substrates. a, TEM image of an array of three InAs nanoribbons on a Si/SiO$_2$ substrate. **b**, Magnified TEM image of an individual ~13-nm-thick InAs nanoribbon on a Si/SiO$_2$ (~50 nm thick) substrate. The nanoribbon is coated with a ZrO$_2$/Ni bilayer (~15 and ~50 nm, respectively), which acts as a top-gate stack for the subsequently fabricated FETs. **c**, HRTEM image showing the single-crystal structure of an InAs nanoribbon with abrupt atomic interfaces with ZrO$_2$ and SiO$_2$ layers on the top and bottom surfaces, respectively. An ~1-nm-thick InAsO$_x$ interfacial layer formed by thermal oxidation and used for surface passivation is clearly evident.

from the measurements. First, the 'off' current monotonically increases with increasing thickness, owing to the reduced electrostatic gate coupling of the back-gate. Second, the 'on' current increases with InAs thickness, owing to the thickness dependence of electron mobility, μ_n. As $L \approx 5\,\mu$m, the devices are effectively operating in the diffusive regime, thereby enabling the direct extraction of the field-effect mobility ($\mu_{n,FE}$) by using the relation $\mu_{n,FE} = g_m(L^2/C_{ox}V_{DS})$, where $g_m = \mathrm{d}I_{DS}/\mathrm{d}V_{GS}|_{V_{DS}}$ is the transconductance, C_{ox} is the gate oxide capacitance, I_{DS} is drain–source current and V_{GS} is gate–source voltage (Supplementary Fig. 5). For this analysis, parasitic resistances were ignored because Ni forms near-ohmic metal contacts[22]. The gate oxide capacitance was estimated from the parallel plate capacitor model $C_{ox} = (\varepsilon A)/d$, where $\varepsilon = 3.9$ and $d = 50$ nm are the dielectric constant and thickness of SiO$_2$, respectively. The effect of quantum capacitance, C_Q, was neglected owing to the relatively thick gate dielectrics used in this study (that is, $C_{ox} \ll C_Q$). Figure 3b shows the peak $\mu_{n,FE}$ as a function of InAs thickness, T_{InAs}. The mobility at first linearly increases with thickness for $T_{InAs} < \sim 18$ nm with a slope of ~221 cm^2 V^{-1} s^{-1} nm^{-1}, beyond which it nearly saturates at $\mu_{n,FE} \approx 5{,}500$ cm^2 V^{-1} s^{-1}. The measured XOI field-effect mobility is close to the reported Hall mobilities for InGaAs (~10,000 cm^2 V^{-1} s^{-1})[10] and InAs (13,200 cm^2 V^{-1} s^{-1})[23] quantum well structures. It should be noted that the Hall mobility is typically higher than the field-effect mobility for any given material, as a number of device and surface state contributions to carrier transport are not accounted for in the Hall effect measurements.

To shed light on the observed mobility trend, the low-field phonon mobility, $\mu_{n,phonon}$, was calculated as $\mu_{n,phonon} = e/(m^*\langle 1/\tau \rangle)$, where e is the electronic charge, and m^* is the effective mass (Supplementary Information). Average scattering rate $\langle 1/\tau \rangle$ is calculated from

$$\langle 1/\tau \rangle = \frac{\int \frac{1}{\tau(E)}\frac{\partial f_0}{\partial E}\mathrm{d}E}{\int \frac{\partial f_0}{\partial E}\mathrm{d}E}$$

where f_0 is the equilibrium Fermi–Dirac distribution function. $\tau(E)$ was calculated using Fermi's golden rule, with the matrix elements of the scattering potentials evaluated in the basis of the nanoribbon eigenfunctions. Both acoustic and optical (including polar) phonon scattering events were considered[24]. The plot of calculated $\mu_{n,phonon}$ versus T_{InAs} is shown in Fig. 3b. For small thicknesses, the mobility increases linearly with the thickness. This behaviour is attributed to the gradual transition of the channel from a two-dimensional to a three-dimensional system as the nanoribbon thickness is increased, with more transport modes (that is, sub-bands) contributing to the current flow. As the thickness increases to a value greater than the Bohr radius of bulk InAs (~34 nm), the electronic structure of the nanoribbons approaches the three-dimensional regime, resulting in a mobility saturation (for $T_{InAs} > \sim 35$ nm) to the well-known bulk value of InAs (~40,000 cm^2 V^{-1} s^{-1})[25]. Whereas the thickness for the onset of saturation closely matches the experiments, there is a discrepancy of 5–10 times in the actual mobility values. This is expected, as the extracted data represent the field-effect mobility, consisting of phonon scattering along with other device contributions (including interface trap states, surface roughness scattering, and vertical-field-induced mobility degradation). Both surface roughness and vertical field (that is, gate field) induce additional carrier scattering events at the surface/interface, while the interface trap states cause the gate-channel coupling efficiency to deteriorate. These effects degrade the extracted g_m and thereby $\mu_{n,FE}$.

To simulate $\mu_{n,FE}$, a full device simulation was performed (Supplementary Information). Using an interface trap density, D_{it}, as the fitting parameter; we obtained $D_{it} = 6 \times 10^{12}$ states cm^{-2} eV^{-1}. The simulated current–voltage (I–V) characteristics of XOI back-gated FETs are shown in Fig. 3a. Clearly, the simulated I–V curves match the experimental data closely for all InAs thicknesses, especially in the on-state. Next, peak $\mu_{n,FE}$ was extracted from the simulation and
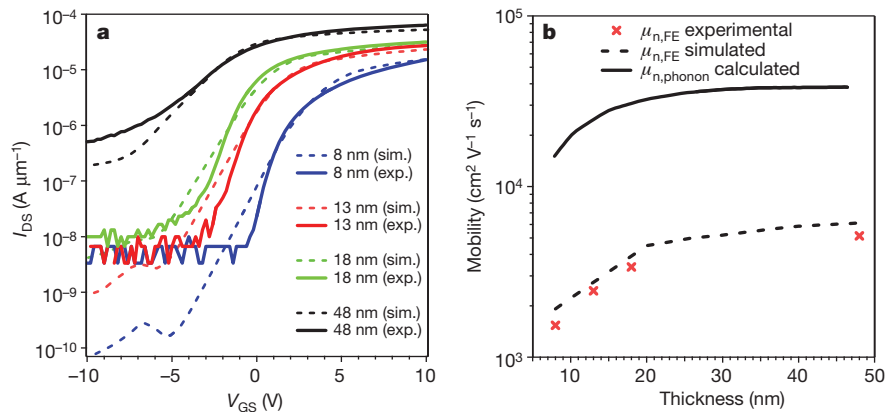
Figure 3 | Back-gated, long-channel InAs XOI FETs. a, Experimental (solid lines) and simulated (dashed lines) I_{DS}–V_{GS} characteristics of back-gated (50-nm SiO$_2$ gate dielectric) XOI FETs at $V_{DS} = 0.1$ V with $L \approx 5$ μm for different InAs nanoribbon thicknesses (8, 13, 18, 48 nm). Each FET consists of a single nanoribbon. **b,** Experimental and simulated peak field-effect electron mobilities ($\mu_{n,FE}$) of InAs nanoribbons as a function of nanoribbon thickness. The calculated phonon mobility ($\mu_{n,phonon}$) is also shown.

plotted as a function of T_{InAs} (Fig. 3b), once again closely matching experimental $\mu_{n,FE}$. The close matching of the experimental and simulated results demonstrate the effectiveness of the XOI platform as a clean and predictable material system for exploring high-performance devices while highlighting the critical role of quantum confinement and surface contributions in the transport properties of InAs, even at relatively large thicknesses. It should be noted that since the ribbon width used in this work is 10–30 times larger than the thickness, there is minimal dependence of the device performance on nanoribbon width (Supplementary Fig. 13), so the structures can be effectively treated as thin films.

In order to explore the performance limits of InAs XOI devices, top-gated FETs with high-dielectric-constant (high-κ) gate insulators and $L \approx 0.5$ μm were fabricated. Briefly, Ni S/D contacts were lithographically patterned on InAs nanoribbons, followed by the atomic layer deposition of ~8-nm-thick ZrO$_2$ ($\varepsilon \approx 20$) as the gate dielectric. A local top-gate (Ni, 50 nm thick), underlapping the S/D electrodes by ~100 nm, was then lithographically patterned. Importantly, thermal oxidation of InAs was found to significantly improve the interfacial properties and FET characteristics (Supplementary Fig. 8). In this regard, before the S/D contact formation, the XOI substrates were first treated with 3% NH$_4$OH to remove the native oxide, followed by the thermal oxidation at 350 °C for 1 min to form an ~1-nm-thick InAsO$_x$ layer (as observed from TEM analysis; Fig. 2c).

Figure 4a shows a typical I_{DS}–V_{GS} characteristic of such a top-gated FET, which consists of an individual ~18-nm-thick InAs nanoribbon with a width of ~320 nm. The XOI FET exhibits a respectable on/off current ratio of 10^4, a subthreshold swing of SS = dV_{GS}/d(logI_{DS}) ≈ 150 mV per decade (Fig. 4a), and a peak g$_m \approx 1.6$ mS μm^{-1} at $V_{DS} = 0.5$ V (Supplementary Fig. 9). The lowest measured SS for our XOI FETs is ~107 mV per decade (Supplementary Fig. 10), as compared to InAs and InGaAs quantum-well FETs in the literature which have exhibited SS values of ~70 and 75 mV per decade, respectively[10,23]. The devices reported here use a relatively thick gate dielectric, which could be scaled down in the future to further improve the gate electrostatic control and the SS characteristics. The single nanoribbon transistor output characteristic is shown in Fig. 4b, delivering an impressive 'on' current of 1.4 mA μm^{-1} at an operating voltage $V_{DD} = V_{DS} = V_{GS} = 1$ V. To further analyse the performance, a full device simulation was carried out. A close match to the experimental data was obtained with fitting parameter $D_{it} = 10^{11}$ states cm^{-2} eV^{-1} (Supplementary Fig. 7), which is a ~60× improvement over devices without any surface treatment (that is, with a native oxide layer). The fitted D_{it} values represent only estimates. Note that while capacitance–voltage (C–V) measurement is conventionally used for D_{it} extraction in Si devices, doing so is rather challenging and prone to a large uncertainty for narrow-bandgap semiconductors, such as InAs (ref. 26). In the future, the development of more accurate techniques for D_{it} measurement in InAs XOI devices is needed. The explored thermal oxidation process for surface passivation is counter-intuitive, as previous work has focused on the removal of surface oxides[7]. We speculate that unlike the native oxide layer, thermal oxidation results in the formation of a dense oxide with minimal dangling bonds. Similar to thermally grown SiO$_2$, the thermal oxide of InAs provides an ideal
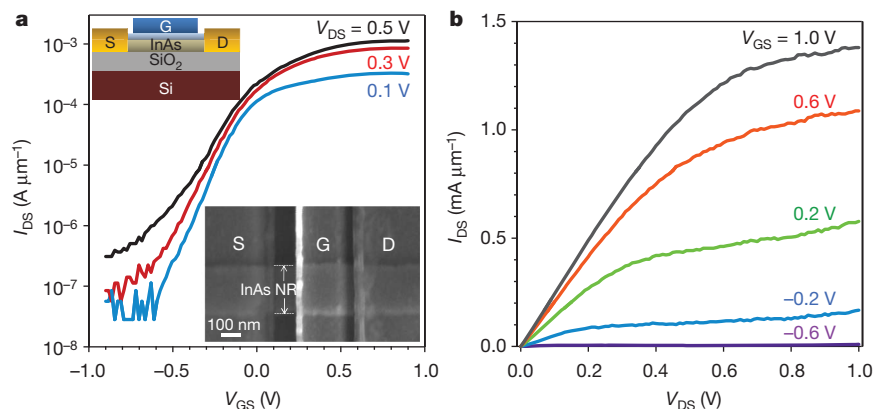
Figure 4 | Top-gated InAs XOI FETs. a, Transfer characteristics of a top-gated InAs XOI FET, consisting of an individual nanoribbon (~18 nm thick) with $L \approx 0.5$ μm and an 8-nm-thick ZrO$_2$ gate dielectric. Inset, device schematic (top) and a representative SEM image (bottom) of a top-gated FET. NR, nanoribbon. **b,** Output characteristics of the same device shown in **a**. nanoribbons were thermally oxidized at 350 °C for 1 min to form ~1-nm-thick interfacial InAsO$_x$ layer for surface passivation of InAs.

and simple surface passivation layer, addressing one of the important challenges in InAs devices.

We have demonstrated a new technology platform and device concept for the integration of ultrathin layers of III–V semiconductors directly on Si substrates, enabling excellent electronic device performance. Although in this work we have focused on InAs as the active channel material, other compound semiconductors could be explored in the future, using a similar scheme. Future research on the scalability of the process for 8-inch and 12-inch wafer processing is needed. We suggest that the direct bonding of Si/SiO$_2$ and III–V wafers, followed by the etch release of the sacrificial layer, might be used in the future to manufacture ultrathin XOI devices on the wafer-scale.

METHODS SUMMARY

Single-crystal InAs thin films (10–100 nm thick) were grown epitaxially on a 60-nm-thick Al$_{0.2}$Ga$_{0.8}$Sb layer on bulk GaSb substrates (Supplementary Fig. 1). Polymethylmethacrylate (PMMA) patterns with a pitch and line-width of ~840 nm and ~350 nm, respectively, were lithographically patterned on the surface of the source substrate. The InAs layer was then pattern etched into nanoribbons using a mixture of citric acid (1 g per ml of water) and hydrogen peroxide (30%) at 1:20 volume ratio, which was chosen for its high selectivity and low resulting InAs edge roughness[27]. To release the InAs nanoribbons from the source substrate, the AlGaSb sacrificial layer was selectively etched by ammonium hydroxide (3% in water) solution for 110 min (ref. 28). Note that the selective etch rate of the AlGaSb layer was high enough not to affect the nanoscale structure of the InAs nanoribbons (Supplementary Fig. 2). Next, an elastomeric polydimethylsiloxane (PDMS) substrate (~2 mm thick) was used to detach the partially released InAs nanoribbons from the GaSb donor substrates and transfer them onto Si/SiO$_2$ (50 nm, thermally grown) receiver substrates by a stamping process (Supplementary Figs 3, 4)[29]. Notably, in this process scheme, the initial epitaxial growth process is used to control the thickness of the transferred InAs nanoribbons, while the lithographically defined PMMA etch mask is used to tune the length and width.

1. Lundstrom, M. Moore's law forever? *Science* **299,** 210–211 (2003).
2. Heyns, M. & Tsai, W. Ultimate scaling of CMOS logic devices with Ge and III–V materials. *Mater. Res. Soc. Bull.* **34,** 485–488 (2009).
3. Theis, T. N. & Solomon, P. M. It's time to reinvent the transistor! *Science* **327,** 1600–1601 (2010).
4. Chau, R., Doyle, B., Datta, S., Kavalieros, J. & Zhang, K. Integrated nanoelectronics for the future. *Nature Mater.* **6,** 810–812 (2007).
5. Javey, A., Guo, J., Wang, W., Lundstrom, M. & Dai, H. Ballistic carbon nanotube transistors. *Nature* **424,** 654–657 (2003).
6. Wong, P. H.-S. Beyond the conventional transistor. *Solid-State Electron.* **49,** 755–762 (2005).
7. Wu, Y. Q., Xu, M., Wang, R. S., Koybasi, O. & Ye, P. Y. High performance deep-submicron inversion-mode InGaAs MOSFETs with maximum G$_m$ exceeding 1.1 mS/um: new HBr pretreatment and channel Engineering. *IEEE IEDM Tech. Digest* **2009,** 323–326 (2009).
8. Bryllert, T., Wernersson, L.-E., Froberg, L. E. & Samuelson, L. Vertical high-mobility wrap-gated InAs nanowire transistor. *IEEE Electron Device Lett.* **27,** 323–325 (2006).
9. Liu, Y. *et al.* in *Fundamentals of III–V Semiconductor MOSFETs* (eds Oktyabrsky, S. & Ye, P.) 31–46 (Springer, 2010).
10. Radosavljevic, M. *et al.* Advanced high-k gate dielectric for high-performance short-channel In$_{0.7}$Ga$_{0.3}$As quantum well field effect transistors on silicon substrate for low power logic applications. *IEEE IEDM Tech. Digest* **2009,** 319–322 (2009).
11. Javorka, P. *et al.* AlGaN/GaN HEMTs on (111) silicon substrates. *IEEE Electron Device Lett.* **23,** 4–6 (2002).
12. Balakrishnan, G. *et al.* Room-temperature optically-pumped GaSb quantum well based VCSEL monolithically grown on Si (100) substrate. *Electron. Lett.* **42,** 350–351 (2006).
13. Yonezu, H. Control of structural defects in group III–V-N alloys grown on Si. *Semicond. Sci. Technol.* **17,** 762–768 (2002).
14. Celler, G. K. & Cristoloveanu, S. Frontiers of silicon-on-insulator. *J. Appl. Phys.* **93,** 4955–4978 (2003).
15. Yablonovitch, E., Hwang, D. M., Gmitter, T. J., Florez, L. T. & Harbison, J. P. Van der Waals bonding of GaAs epitaxial liftoff films onto arbitrary substrates. *Appl. Phys. Lett.* **56,** 2419–2421 (1990).
16. Kim, D.-H. *et al.* Ultrathin silicon circuits with strain-isolation layers and mesh layouts for high-performance electronics on fabric, vinyl, leather, and paper. *Adv. Mater.* **21,** 3703–3707 (2009).
17. Melosh, N. *et al.* Ultrahigh density nanowire lattices and circuits. *Science* **300,** 112–115 (2003).
18. Yokoyama, M. *et al.* III–V-semiconductor-on-insulator n-channel metal-insulator-semiconductor field-effect transistors with buried Al$_2$O$_3$ layers and sulfur passivation: Reduction in carrier scattering at the bottom interface. *Appl. Phys. Lett.* **96,** 142106 (2010).
19. Yuan, H.-C. & Ma, Z. Microwave thin-film transistors using Si nanomembranes on flexible polymer substrate. *Appl. Phys. Lett.* **89,** 212105 (2006).
20. Kim, D.-H. *et al.* Stretchable and foldable silicon integrated circuits. *Science* **320,** 507–511 (2008).
21. Yoon, J. *et al.* GaAs photovoltaics and optoelectronics using releasable multilayer epitaxial assemblies. *Nature* **465,** 329–333 (2010).
22. Chueh, Y.-L. *et al.* Formation and characterization of Ni$_x$InAs/InAs nanowire heterostructures by solid source reaction. *Nano Lett.* **8,** 4528–4533 (2008).
23. Kim, D.-H. *et al.* Scalability of sub-100 nm InAs HEMTs on InP substrate for future logic applications. *IEEE Trans. Electron. Dev.* **57,** 1504–1511 (2010).
24. Lundstrom, M. *Fundamentals of Carrier Transport* 54–118 (Cambridge Univ. Press, 2000).
25. Mikhailova, M. P. in *Handbook Series of Semiconductor Parameters* Vol. 1, *Elementary Semiconductors and A3B5 Compounds Si, Ge C, GaAs, GaP, GaSb InAs, InP, InSb* (eds Levinshtein, M., Rumyantsev, S. & Shur, M.) 31–46 (World Scientific, 1996).
26. Martens, K. *et al.* On the correct extraction of interface trap density of MOS devices with high-mobility semiconductor substrates. *IEEE Trans. Electron. Dev.* **55,** 547–556 (2008).
27. DeSalvo, G. C., Kaspi, R. & Bozada, C. A. Citric acid etching of GaAs$_{1-x}$Sb$_x$, Al$_{0.5}$Ga$_{0.5}$Sb, and InAs for heterostructure device fabrication. *J. Electrochem. Soc.* **141,** 3526–3531 (1994).
28. Yoh, K., Kiyomi, K., Nishida, A. & Inoue, M. Indium arsenide quantum wires fabricated by electron beam lithography and wet-chemical etching. *Jpn. J. Appl. Phys.* **31,** 4515–4519 (1992).
29. Meitl, M. A. *et al.* Transfer printing by kinetic control of adhesion to an elastomeric stamp. *Nature Mater.* **5,** 33–38 (2006).

LETTER

Early oxygenation of the terrestrial environment during the Mesoproterozoic

John Parnell[1]*, Adrian J. Boyce[2], Darren Mark[2], Stephen Bowden[1] & Sam Spinks[1]

Geochemical data from ancient sedimentary successions provide evidence for the progressive evolution of Earth's atmosphere and oceans[1–7]. Key stages in increasing oxygenation are postulated for the Palaeoproterozoic era (~2.3 billion years ago, Gyr ago) and the late Proterozoic eon (about 0.8 Gyr ago), with the latter implicated in the subsequent metazoan evolutionary expansion[8]. In support of this rise in oxygen concentrations, a large database[1–3,9] shows a marked change in the bacterially mediated fractionation of seawater sulphate to sulphide of $\Delta^{34}S < 25‰$ before 1 Gyr to $\geq 50‰$ after 0.64 Gyr. This change in $\Delta^{34}S$ has been interpreted to represent the evolution from single-step bacterial sulphate reduction to a combination of bacterial sulphate reduction and sulphide oxidation, largely bacterially mediated[3,7,9]. This evolution is seen as marking the rise in atmospheric oxygen concentrations and the evolution of non-photosynthetic sulphide-oxidizing bacteria[3,7,10]. Here we report $\Delta^{34}S$ values exceeding 50‰ from a terrestrial Mesoproterozoic (1.18 Gyr old) succession in Scotland, a time period that is at present poorly characterized. This level of fractionation implies disproportionation in the sulphur cycle, probably involving sulphide-oxidizing bacteria, that is not evident from $\Delta^{34}S$ data in the marine record[1–3,9]. Disproportionation in both red beds and lacustrine black shales at our study site suggests that the Mesoproterozoic terrestrial environment was sufficiently oxygenated to support a biota that was adapted to an oxygen-rich atmosphere, but had also penetrated into subsurface sediment.

The isotope fractionation between sulphate and sulphide ($\Delta^{34}S = \delta^{34}S_{sulphate} - \delta^{34}S_{sulphide}$) is a valuable measure of microbial processing

of sulphur species[9]. Bacterial sulphate reduction alone induces a fractionation up to ~45‰ but a higher limit of about 75‰ is suggested by recent models that involve multiple intermediate stages in the reduction process[10] or extreme single-step isotopic fractionation[11]. However, typical $\Delta^{34}S$ during the Proterozoic is 20–25‰ (Fig. 1a and refs 3, 12). The change has been interpreted to mark the evolution of non-photosynthetic sulphide-oxidizing bacteria (contributing to repeated oxidation-disproportionation cycles, which allow greater isotopic fractionation). Regardless of the mechanistic detail, the extended $^{34}S/^{32}S$ fractionation indicates a step change in sulphur cycling[7]. The first expression of this transition evident in $\delta^{34}S$ data was interpreted as about 0.8–0.7 Gyr ago[1,3,12], although a different evaluation suggests about 0.58 Gyr ago[13]. However, the use of $\Delta^{33}S$ data[14] suggests that disproportionation had already developed by the Mesoproterozoic[14], becoming progressively more important to the global sulphur cycle from 1.45 to 1.2 Gyr ago. This record has been argued to be sensitive to microbial disproportionation at levels of $^{34}S/^{32}S$ fractionation that could be explained by sulphate reduction alone, and hence would not be evident from $\delta^{34}S$ data. There is also evidence for an oceanic oxygenation event at around 1.3 Gyr ago, consistent with a carbon isotope excursion at 1.3–1.25 Gyr ago that indicates organic carbon burial[12] and enhanced precipitation of marine gypsum[15]. Despite evidence for increased ocean sulphate concentrations, the bacteriogenic fractionation of sulphate to sulphide ($\Delta^{34}S$) remained consistently below 25‰, indicating that either the model[1] was incorrect, or that sulphide-oxidizing bacteria had not evolved at this stage[15]. At this early stage, it was suggested that disproportionation occurred predominantly
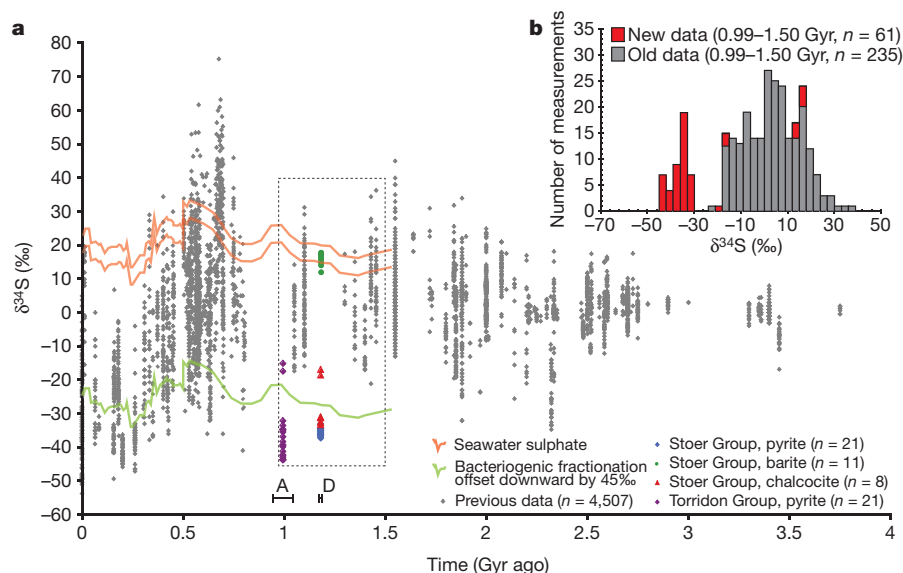
Figure 1 | Compilation of published and new sulphur isotope compositions for sedimentary pyrite. **a,** Published compositions ($n = 4,507$) from ref. 1. Time markers: A, Diabaig Formation whole-rock Rb–Sr isochron age of 0.994 ± 0.048 Gyr (ref. 19); D, Stac Fada Member Ar/Ar age for the Stac Fada

Member of 1.178 ± 0.009 Gyr ago. **b,** Histogram of 500-Myr subset from **a**, showing that the new data ($n = 61$) define a second $\delta^{34}S$ population within the 1 to 1.5 Gyr ago time window relative to previous measurements ($n = 235$).

[1]School of Geosciences, University of Aberdeen, Aberdeen AB24 3UE, UK. [2]Scottish Universities Environmental Research Centre, East Kilbride, Glasgow G75 0QF, UK.

in surface environments, where local oxidative processes produced the intermediate sulphur compounds that allow disproportionation[14]. Evidence of Mesoproterozoic disproportionation in the $\delta^{34}S$ record is therefore most likely to be found in terrestrial environments close to 1.2 Gyr old.

The late Mesoproterozoic Stoer Group rocks of Scotland (Figs 2 and 3 and Supplementary Information) are especially suitable for investigating the record of disproportionation. They are about 1.2 Gyr old, exceptionally well preserved, and deposited in a range of sulphate-rich terrestrial environments in a non-marine basin[16]. They contain stromatolitic limestones and microfossils[16], as well as sedimentary fabrics indicative of an active biota[17]. Although most lakes have low sulphate concentrations and sulphur isotope fractionation[18], high sulphate concentrations in the Stoer Group engendered by a closed basin hydrological regime[16] would have allowed isotopic fractionation to develop to maximum potential[9]. We sought evidence of this fractionation in $\Delta^{34}S$ values from Stoer Group sulphides and sulphates.

The Stoer Group has previously been dated at 1.2–1.15 Gyr old, through Pb–Pb dating of limestone[19] and palaeomagnetic data[20]. It is unconformably overlain by the approximately 1.0-Gyr-old Torridon Group[16] (Fig. 3). The Stoer Group rocks are broadly divisible into a predominant red bed facies, including alluvial fan conglomerates, fluviatile and aeolian sandstones, and shallow lacustrine mudrocks, and a black shale facies representing permanent, anoxic lacustrine conditions[16]. The black shale facies overlies a horizon interpreted as an impact ejecta deposit (the Stac Fada Member; Fig. 3 and ref. 21), which potentially created the topography for establishment of the lake. Both red beds and black shales contain pseudomorphs after gypsum[16]. A sulphate progenitor is indicated by barite within the black shale pseudomorphs (Fig. 4), and barite veinlets cutting the

Figure 3 | Summary stratigraphic section of Stoer and Torridon groups. **a**, Section with relevant geochronologial information. **b**, Blown up from **a**. Lithostratigraphic log of Stoer Group at Stoer Bay (after ref. 16) showing sampling horizons, and time markers: A, Diabaig Formation whole-rock Rb-Sr isochron age of 994 ± 48 Myr (ref. 19); B, Stoer Group limestone Pb–Pb isochron age of 1,199 ± 70 Myr (ref. 19); C, Stoer Group palaeomagnetic age of 1,180 Myr (ref. 20) and D, Stac Fada Member Ar/Ar age for the Stac Fada Member of 1.178 ± 0.009 Gyr ago.

pseudomorph-bearing red mudstones. Hence we take the barite to represent the original sulphate, as in previous studies[22].

The occurrence of sulphate evaporites in the sequence is also reflected in the widespread precipitation of diagenetic sulphides in the black shale facies (Fig. 4). Millimetre-scale pyrite nodules coalesced parallel to bedding. The copper sulphide chalcocite occurs as diagenetic nodules in sandstones at two localities[16]. In the oxidizing conditions represented by terrestrial red beds, copper is mobile and is locally precipitated as sulphides, with microbial mediation[23]. The diagenetic nodules in the red beds and black shale formed before complete compaction of the host rock, as evidenced by deflection of laminae around the nodules, which would be expected within 50 million years (Myr) of the 1.2 Gyr depositional age, and unequivocally before the deposition about 1.0 Gyr ago of the overlying Torridon Group, which followed uplift of the strongly lithified Stoer Group[16].

We determined a precise $^{40}Ar/^{39}Ar$ age of 1.178 ± 0.009 Gyr ago (2σ, 92% ^{39}Ar, $n = 10$, mean square weighted deviation, MSWD = 0.9) by dating authigenic K-feldspar precipitated at high temperatures during cooling of the impact ejecta deposit in Stoer Bay. The emplacement of the Stac Fada Member and the precipitation of the K-feldspar were synchronous (Supplementary Information), and provide a chronostratigraphic marker that is consistent with, but more precise than, previous estimations[19,20] (Fig. 3).

Measurements of sulphur isotope composition were focused on the type section at Stoer Bay (Fig. 2 and Supplementary Information), which includes both red bed and black shale facies. 11 sulphates and 26 sulphides were analysed. The mean sulphur isotope composition of these continental sulphates is 16.2 ± 1.8‰ ($n = 11$), which are significantly lighter than the 22‰ to 32‰ reported for 1.2-Gyr-old marine gypsum[15], but consistent with most models for 1.2-Gyr-old seawater sulphate[1,3] (Fig. 1a). It is thus reasonable to suggest that the homogeneous sulphate $\delta^{34}S$ is representative of Stoer Group lake sulphate. The sulphides, in contrast, are strongly enriched in ^{32}S, showing very negative $\delta^{34}S$. The red-bed-hosted chalcocite has a range of compositions from −33.2 to

Figure 2 | Map of NW Scotland showing outcrop of Stoer Group and sample localities. Four localities for black/grey shales are all mineralized by sulphides. Diagenetic sulphides occur at BS (Bay of Stoer) and CM (Clashmore). Crosscutting sulphide veins (not analysed) occur at AB (Achiltibuie) and SP (Stattic Point). Sulphates (barite) occur at Clachtoll (CT) and Sgeir na Traghad (ST). Base map after ref. 16. Inset shows location within mainland UK.

Figure 4 | Photomicrographs of Stoer Group sulphide/sulphate occurrences in shale host. a, Pseudomorph after gypsum in shale, representing primary sulphate. **b,** Backscattered electron image of barite (bright) with quartz and calcite forming pseudomorph after gypsum (rhombic shape outlined for clarity). **c,** Pyrite (bright) nodule. We note that pyrite is pre-compactional and therefore diagenetic.

−16.8 ‰, indicating a maximum fractionation of 51.2‰. The black-shale-hosted sulphide has a range from −33.0 to −37.2‰ (mean −35.4 ± 1.0‰; $n = 21$). The maximum fractionation from the sulphate is 55.2‰. These values include, by far, the lightest values of any Mesoproterozoic sulphide. Adopting mean values for barite, pyrite and chalcocite, these data yield mean $\Delta^{34}S$ values of 51.6‰ and 44.5‰ for the pyrite and chalcocite, respectively. A large database of sulphide compositions from younger rocks in the region[24] (range −21‰ to +9‰; $n = 79$) is distinctly heavier, and confirms that the Mesoproterozoic sulphides have not been contaminated by later events.

These large isotopic fractionations are well in excess of the 20‰ or less due to non-biological kinetic isotopic fractionation[25], and both the extent and the distribution of values (Fig. 1b) are typical features of bacterial sulphate reduction. In the Proterozoic database of ref. 1, the Stoer Group (about 1.18 Gyr old) is within the time period characterized by microbial sulphate reduction only (Fig. 1a), but shows values much lighter than the existing database for this time period, and instead similar to the levels of fractionation recorded in younger samples with a proposed greater level of oxidation. No analyses before 0.86 Gyr ago ($n = 383$) have $\Delta^{34}S$ of 45‰ or greater compared to seawater sulphate: indeed, only seven values (less than 2% of total) have $\Delta^{34}S$ of over 35‰. There is no doubt that $\Delta^{34}S$ values for the Stoer Group are much greater than was previously recognized in rocks of this age (Fig. 1b), and that their levels of isotopic fractionation are typical of modern, open-system bacterial sulphate-reducing systems. Yet the Stoer Group is over 300 Myr older than the oldest published data showing modern levels of fractionation and, by implication, oxidation. This interval from 1.18 to 0.86 Gyr ago has relatively little data. To test whether the Stoer Group

data represents the commencement of a trend of isotopically light data, or is an anomalous exception, data was also collected from diagenetic pyrite in lacustrine siltstones of the ~1.0-Gyr-old Diabaig Formation at the base of the younger Torridon Group in the same region (Fig. 2). The Torridon Group pyrite has compositions (mean −30.1 ± 17.3‰; $n = 25$) (Fig. 1b) indicating a continuity of the isotopically light trend in the terrestrial record identified here from about 1.2 Gyr ago.

The large $\Delta^{34}S$ exhibited in both the red beds and black shales of the Stoer Group indicates that availability of carbon was not a controlling factor in microbial activity, as it is likely to have been in normal, low-sulphate, terrestrial environments[18]. Ocean water at 1.2 Gyr ago is assumed to have had a sulphate concentration an order of magnitude lower than today. Such low marine sulphate concentrations have been proposed to explain the limitations of isotopic fractionation ($\Delta^{34}S$) to mostly <35‰ in the Mesoproterozoic[13], as an alternative to the hypothesis that low oxygen concentration in the atmosphere restricted fractionation to microbial sulphate reduction. In contrast, it was argued that if 1.2 Gyr ago oceans had sulphate concentrations similar to today's oceans, then the lack of S isotopic fractionation >25‰ is a sign that sulphide-oxidizing bacteria had not evolved[1,15]. Our data indicate that the Stoer Group waters were not sulphate-limited, and isotopic fractionation was also not limited by oxygen level. The Stoer Group data show clear evidence of large bacteriogenic S isotope fractionation, and therefore disproportionation, probably involving sulphide-oxidizing bacteria. The data imply high oxygenicity in Mesoproterozoic surface environments, while coeval marine environments experienced only limited oxygenation[3–5]. The resulting gradient in redox conditions would be conducive to the development of distinct ecological systems, in which disproportionation evolved at the surface. The evolution of such a range of ecosystems has been predicted[26,27], but the Stoer Group data provides a time constraint which shows that this important step towards the colonization of sediment by a complex biota had arisen by the Mesoproterozoic.

METHODS SUMMARY

Sulphides were separated by heavy liquid and hand picking; sulphide and sulphate samples were then analysed by standard mass spectrometry techniques[28,29]. In addition to the conventional analysis, in situ laser combustion of sulphides was carried out on polished blocks, following the technique outlined in ref. 30. Experiments show that laser combustion results in a small, but predictable fractionation of the sulphur isotope composition $\delta^{34}S$ values of the resulting SO_2 gas compared to the mineral $\delta^{34}S$, and the raw data are thus corrected ($\delta^{34}S_{pyrite} = \delta^{34}S_{SO2}$ laser + 0.8‰; ref. 30). All SO_2 gases were analysed on a VG Isotech SIRA II mass spectrometer. Reproducibility, and hence uncertainty in values, for laser combustion and conventional analysis, is around ±0.3‰ (ref. 30). The standards used were international standards NBS-123 and IAEA-S-3 (supplied by the IAEA), and the SUERC laboratory standard CP-1. These gave $\delta^{34}S$ values of +17.1‰, −31.5‰ and −4.6‰ respectively, with 1σ reproducibility, based on repeat analyses of the standards, better than ±0.2‰. Data are reported in $\delta^{34}S$ notation as per mil (‰) variations from the V-CDT standard.

1. Canfield, D. E. & Teske, A. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* **382,** 127–132 (1996).
2. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396,** 450–453 (1998).
3. Canfield, D. E. The evolution of the Earth surface sulfur reservoir. *Am. J. Sci.* **304,** 839–861 (2004).
4. Bekker, A. *et al.* Dating the rise of atmospheric oxygen. *Nature* **427,** 117–120 (2004).
5. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444,** 744–747 (2006).
6. Scott, C. *et al.* Tracing the stepwise oxygenation of the Proterozoic ocean. *Nature* **452,** 456–459 (2008).
7. Lyons, T. W. & Gill, B. C. Ancient sulfur cycling and oxygenation of the early biosphere. *Elements* **6,** 93–99 (2010).
8. Runnegar, B. Precambrian oxygen levels estimated from the biochemistry and physiology of early eukaryotes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97,** 97–111 (1991).
9. Canfield, D. E. & Raiswell, R. The evolution of the sulfur cycle. *Am. J. Sci.* **299,** 697–723 (1999).

10. Johnston, D. T., Farquhar, J. & Canfield, D. E. Sulfur isotope insights into microbial sulphate reduction: when microbes meet models. *Geochim. Cosmochim. Acta* **71,** 3929–3947 (2007).
11. Wortmann, U., Bernasconi, S. M. & Böttcher, M. E. Hypersulfidic deep biosphere indicates extreme sulfur isotope fractionation during single-step microbial sulfate reduction. *Geology* **29,** 647–650 (2001).
12. Lyons, T. C., Kah, L. C. & Gellatly, A. M. in *The Precambrian Earth: Tempos and Events* (eds Eriksson, P. G., Altermann, W., Nelson, D. R., Mueller, W. U. & Catuneanu, O.) 421–440 (Elsevier, 2004).
13. Hurtgen, M. T., Arthur, M. A. & Halverson, G. P. Neoproterozoic sulfur isotopes, the evolution of microbial sulfur species, and the burial efficiency of sulphide as sedimentary pyrite. *Geology* **33,** 41–44 (2005).
14. Johnston, D. T. *et al.* Active microbial sulfur disproportionation in the Mesoproterozoic. *Science* **310,** 1477–1479 (2005).
15. Kah, L. C., Lyons, T. W. & Chesley, J. T. Geochemistry of a 1.2 Ga carbonate-evaporite succession, northern Baffin and Bylot Islands: implications for Mesoproterozoic marine evolution. *Precambr. Res.* **111,** 203–234 (2001).
16. Stewart, A. D. The later Proterozoic Torridonian rocks of Scotland: their sedimentology, geochemistry and origin. *Geol. Soc. Mem.* No. **24** (2002).
17. Prave, A. R. Life on land in the Proterozoic: evidence from the Torridonian rocks of northwest Scotland. *Geology* **30,** 811–814 (2002).
18. Fry, B., Giblin, A., Dornblaser, M. & Peterson, B. in *Geochemical Transformations of Sedimentary Sulfur* (eds Vairavamurthy, A. & Schoonens, M. A. A.) 397–410 (American Chemical Society, 1995).
19. Turnbull, M. J. M., Whitehouse, M. J. & Moorbath, S. New isotopic age determinations for the Torridonian, NW Scotland. *J. Geol. Soc. Lond.* **153,** 955–964 (1996).
20. Darabi, M. H. & Piper, J. D. A. Palaeomagnetism of the (Late Mesoproterozoic) Stoer Group, Northwest Scotland: implications for diagenesis, age and relationship to the Grenville Orogeny. *Geol. Mag.* **141,** 15–39 (2004).
21. Amor, K., Hesselbo, S. P., Porcelli, D., Thackrey, S. & Parnell, J. A Precambrian proximal ejecta blanket from Scotland. *Geology* **36,** 303–306 (2008).
22. Strauss, H. The sulfur isotopic record of Precambrian sulfates: new data and a critical evaluation of the existing record. *Precambr. Res.* **63,** 225–246 (1993).
23. Sillitoe, R. H., Folk, R. L. & Saric, N. Bacteria as mediators of copper sulfide enrichment during weathering. *Science* **272,** 1153–1155 (1996).
24. Lowry, D., Boyce, A. J., Fallick, A. E., Stephens, W. E. & Grassineau, N. V. Terrane and basement discrimination in northern Britain using sulphur isotopes and mineralogy of ore deposits. *Geol. Soc. Lond. Spec. Publ.* **248,** 133–151 (2005).
25. Machel, H. G. Bacterial and thermochemical sulphate reduction in diagenetic settings — old and new insights. *Sedim. Geol.* **140,** 143–175 (2001).
26. Johnston, D. T., Wolfe-Simon, F., Pearson, A. & Knoll, A. H. Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc. Natl Acad. Sci. USA* **106,** 16925–16929 (2009).
27. Poulton, S. W., Fralick, P. W. & Canfield, D. E. Spatial variability in oceanic redox structure 1.8 billion years ago. *Nature Geosci.* **3,** 486–490 (2010).
28. Robinson, B. W. & Kusakabe, M. Quantitative preparation of sulfur dioxide for $^{34}S/^{32}S$ analyses from sulphides by combustion with cuprous oxide. *Anal. Chem.* **47,** 1179–1181 (1975).
29. Coleman, M. L. & Moore, M. P. Direct reduction of sulphates to sulphur dioxide for isotopic analysis. *Anal. Chem.* **28,** 199–260 (1978).
30. Wagner, T., Boyce, A. J. & Fallick, A. E. Laser combustion analysis of $\delta^{34}S$ of sulfosalt minerals: determination of the fractionation systematics and some crystal-chemical considerations. *Geochim. Cosmochim. Acta* **66,** 2855–2863 (2002).

LETTER

Sequential faulting explains the asymmetry and extension discrepancy of conjugate margins

César R. Ranero[1]* & Marta Pérez–Gussinyé[2]*

During early extension, cold continental lithosphere thins and subsides, creating rift basins. If extension continues to final break-up, the split and greatly thinned plates subside deep below sea level to form a conjugate pair of rifted margins. Although basins and margins are ubiquitous structures, the deformation processes leading from moderately extended basins to highly stretched margins are unclear, as studies consistently report that crustal thinning is greater than extension caused by brittle faulting[1–4]. This extension discrepancy might arise from differential stretching of brittle and ductile crustal layers[2], but that does not readily explain the typical asymmetric structure of conjugate margins[5,6]—in cross-section, one margin displays gradual thinning accompanied by large faults, and the conjugate margin displays abrupt thinning but smaller-scale faulting[5]. Whole-crust detachments, active from early in the rifting, could in theory create both thinning and asymmetry[1], but are mechanically problematical. Furthermore, the extension discrepancy occurs at both conjugate margins, leading to the apparent contradiction that both seem to be upper plates to a detachment fault[7,8]. Alternative models propose that much brittle extension is undetected because of seismic imaging limitations caused either by subseismic-resolution faulting[9], invisible deformation along top-basement 100-km-scale detachments[8] or the structural complexity of cross-cutting arrays of faults[3]. Here we use depth-migrated seismic images to accurately measure fault extension and compare it with crustal thinning. The observations are used to create a balanced kinematic model of rifting that resolves the extension discrepancy by producing both fault-controlled crustal thinning which progresses from a rift basin to the asymmetric structure, and extreme thinning of conjugate rifted margins. Contrary to current wisdom, the observations support the idea that thinning is to a first degree explained by simple Andersonian faulting that is unambiguously visible in seismic data.

Our model is based on data from the west Iberia margin (WIM) and its conjugate Newfoundland margin (NFM) that are archetypical for highly extended non-volcanic margins and exemplify rifting of initially cold lithosphere[8,10] (Fig. 1). We analysed multibeam bathymetry[11], new pre-stack depth-migrated images of the WIM, and re-processed and pre-stack depth-migrated data from NFM. Bathymetry shows large fault-bounded north–south striking ridges (Fig. 1b) indicating that large-scale structures imaged on seismic line IAM11 are essentially two dimensional and, thus, representative of the rifting process (Fig. 2).

The transition from the ~20-km-thick continental crust of Galicia Bank to mantle exhumation involves seven major faults (Fig. 2). The fault blocks have relatively simple geometries with clear syntectonic strata overlying pre-rift (or earliest syn-rift) sediment, indicating that large fault blocks have not been dismembered by superimposed episodes of large-scale faulting as previously suggested[3]. The large faults f1–f6, with 4–6-km throws, have angular relationships to pre-rift strata which indicate that they formed at ~65–55° angles and rotated to the current ~42–28° (Fig. 2). Those dips differ by only a few degrees from true dips
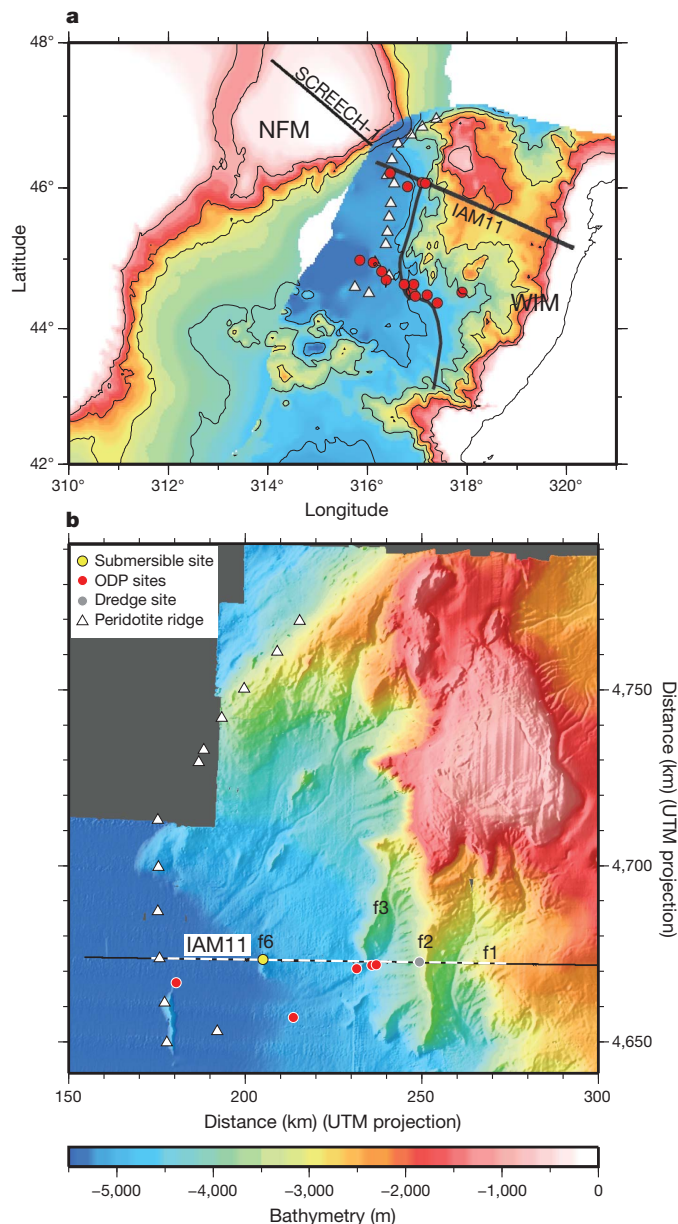
Figure 1 | WIM and NFM. a, Reconstruction at chron M0[29]. Black lines are seismic profiles. IAM11 has been processed and pre-stack depth-migrated, and SCREECH-1 (originally published by ref. 5) has been reprocessed and pre-stack depth-migrated with refined velocities. The black curved line marks oceanward extent of continental crust at the WIM. **b**, Multibeam bathymetry of the deep Galicia margin[11]. The IAM11 segment in Fig. 2 is marked by white dashes. Scarps f1, f2, f3 and f6 correspond to faults on IAM11 (Fig. 2a). ODP, Ocean Drilling Program. UTM, universal transverse Mercator.

[1]ICREA at CSIC, Barcelona Center for Subsurface Imaging, Instituto de Ciencias del Mar, CSIC, Passeig Marítim de la Barcelona 37-49, 08003, Barcelona, Spain. [2]Department of Earth Sciences, Royal Holloway, University of London, Egham TW20 0BD, UK.
*These authors contributed equally to this work.

for f3, where the seismic profile is not exactly perpendicular to the fault plane (Fig. 1). The fault angles are compatible with Andersonian mechanics, which allows slip until an angle of ~30° is reached.

Bathymetry shows that WIM faults are arcuate in plan view, implying that the cross-sectional dimensions of imaged fault blocks are influenced by profile location (Fig. 1b). Despite this, the trend is for fault-block thickness generally to decrease from B1 to B7, from about 18 to 5 km. Studies of normal-fault populations in nature, and numerical and analogue models, establish that fault-block dimensions decrease with decreasing brittle-layer thickness[12,13]. Furthermore, numerical models of cold-lithosphere rifting[14] and pressure–temperature–time paths of lower crustal rocks[15,16] indicate that the lower crust progressively becomes brittle as the crust thins. Thus, the decrease in dimensions from block B1 to block B7 indicates that at the time of fault activity, the brittle layer—which may have included most or all of the crust—must have decreased in thickness from fault f1 to f7 (Fig. 2), and included lower crustal rocks.

The age of faulting at the WIM is debated. Stratigraphic interpretation of 1980s seismic images indicated a rifting ~25 million years (Myr) long[17]. But reinterpretation of the images concluded that syn-rift sediment could not be objectively identified[18], and faulting might have lasted <5 Myr. Superior modern images have permitted us a better identification of pre- and syn-rift strata (Methods and Supplementary

Fig. 1). Sedimentation during f2 activity is older than sedimentation to f4, indicating that movement along faults was not coeval (Fig. 2). Diachronous-fault activity has also been interpreted in the southern segment of the WIM[19].

The above observations may be explained if faults f1–f7 were formed successively and functioned sequentially through time, with each fault causing concomitant crustal thinning. For example, f3 formed on crust that had been previously cut and thinned by fault f2, and f3 thus bounds a smaller crustal block than f2. In this manner, successive f1–f7 fault activity produces gradually smaller B1–B7 crustal blocks (Fig. 2). Had faults f1–f7 functioned synchronously, they would have cut sections of crust of similar thicknesses and bound blocks of similar dimensions, which is not the case.

Comparison of structures on IAM11 with conjugate NFM shows that rifting has created a regional-scale asymmetric configuration. In NFM, Moho is steeper, crustal thinning abrupt, faulting comparatively subdued, and pre-rift and syn-rift sediment less differentiated (Fig. 3). However, coeval tectonic processes must have configured WIM and NFM architecture.

We have constructed a balanced kinematic model consistent with our observations of conjugate, highly extended margins, and with observations of the preceding evolution of faulting in less extended basins. Seismic images provide crustal structure, so our model does not
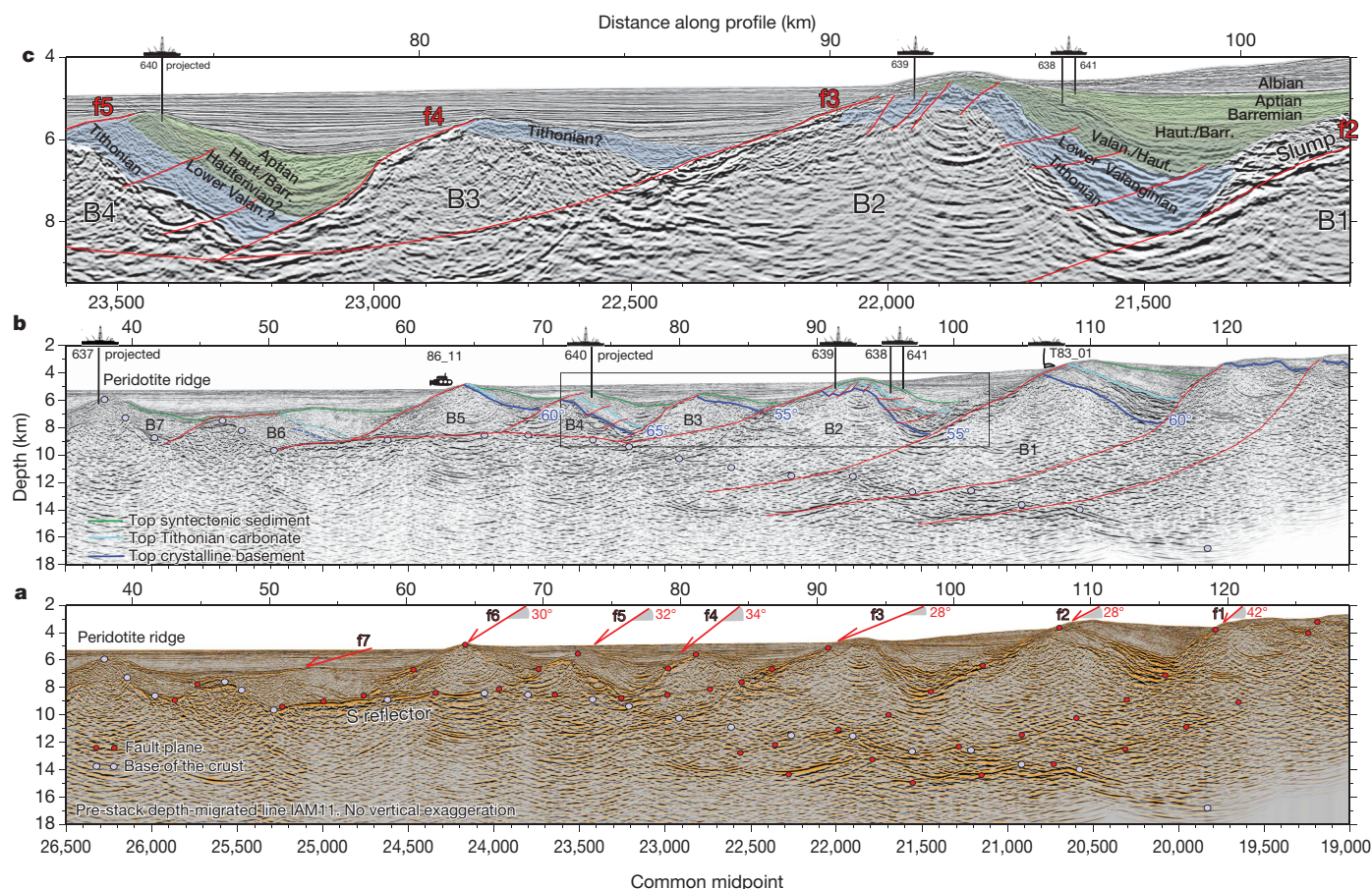
Figure 2 | Pre-stack depth-migrated line IAM11. a, Arrows and numbers indicate the average dips of the block-bounding fault segments exhumed during rifting. Red circles delineate reflections in crust and mantle interpreted as faults, which project up-dip to the exhumed block-bounding fault segments. Grey circles mark the crust base, constrained with wide-angle seismic data[30]. **b**, Tectonic and stratigraphic interpretation calibrated with ODP sites 637–641, dive 86_11 and dredge T83_01 (ref. 17). Faults bounding blocks B1–B5 possibly formed at 55°–65° (in blue), as inferred from their angular relationship to Jurassic strata originally deposited subhorizontally. **c**, Magnified view of box in **b**, showing geometry and calibrated ages of sediment units. At block B2, four

units overlie pre-rift carbonates. A Valanginian/Hauterivian unit, drilled at sites 638 and 639, was deposited during the main activity of fault f2. Younger Hauterivian/Barremian, Barremian/Aptian and Albian/recent units drilled at sites 638 and 641 are post-tectonic. Block B3 is imaged at its edge and sediment units are too thin to be resolved. At site 640, in block B4, Barremian/Aptian sediment was found as part of the section of syn-rift strata. Thus, the activity of fault f2 stopped during Hauterivian time, possibly before the activity of fault f4 began, during Barremian time, and several million years before the end of f4 activity, during late Aptian or early Albian time.
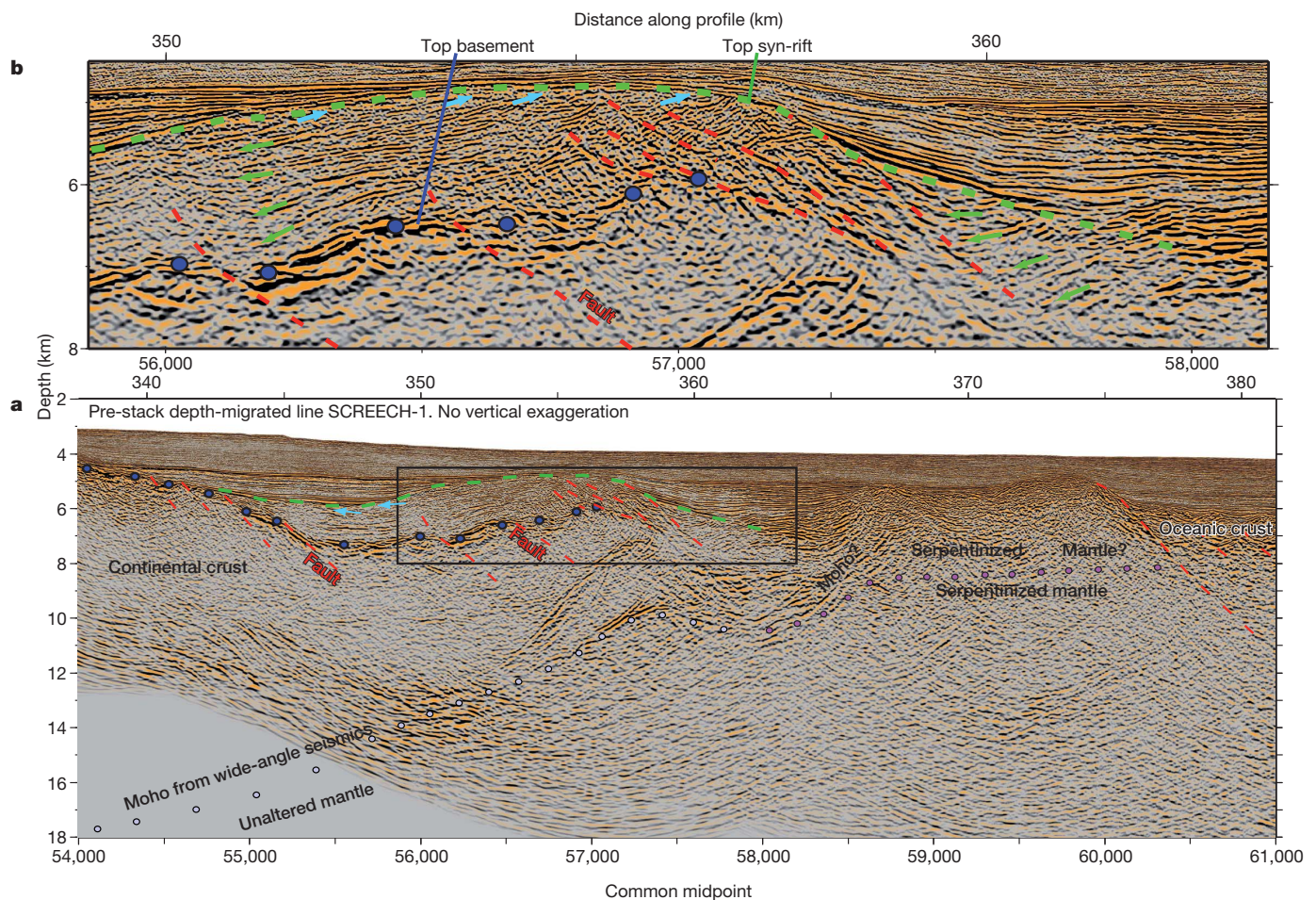
Figure 3 | Interpretation of a pre-stack depth-migrated segment of line SCREECH-1. a, The ~40-km-long segment illustrates the transition from continental to oceanic crust[5], showing the extreme structural asymmetry relative to the conjugated ~80-km-long section on IAM11 (Fig. 2). **b**, Magnified view of the fault block marked by the box in **a**. Green arrows indicate strata fanning caused by syntectonic deposition.

consider lithospheric-mantle deformation. For basins, we use observations from regions where faulting is constrained in space and time, particularly the North Sea[20,21] and the Gulf of Suez[22], and from other well-studied regions such as the Gulf of Corinth[23]. Our model differentiates between two stages, a rift basin stage (Fig. 4a–c) and a subsequent rifted margin stage (Fig. 4d–k), both of which consider the evolving rheology of extending crust. Model constraints are discussed in Methods and Supplementary Fig. 3 shows a stepwise construction.

Rift basin studies show that early extension is distributed across numerous disconnected, short faults with no dominant inward or outward sense of dip[13,20]. In the WIM, this corresponds to shallow-water carbonate sedimentation found by drilling and dredging (Fig. 2). In this phase, faults rotate a few degrees (Fig. 4a). With increasing extension, faults grow laterally and link. Continued deformation focuses in a few active, larger faults. Simultaneously, outward-dipping faults halt, and strain concentrates in inward-dipping faults near the future basin centre, narrowing the area of active faulting[20–22]. Faults in less thinned rift flanks become inactive (Fig. 4b). During this phase, the ductile lower crust[14] deforms by localized thinning below the faulting region (Supplementary Fig. 2). Evidence for this comes from numerical models of cold lithosphere showing that the lower crust cannot flow long distances, but deforms in compliance with the upper crust[24,25]. During the last phase of the rift basin stage, strain migrates to a single fault defining the basin centre[20]. Whereas other faults stop at higher angles, a last fault (fc in Fig. 4c) works until it is no longer mechanically possible, and eventually all faults are locked.

However, stresses continue pulling plates apart and strain must be accommodated by new deformation. We propose this new phase as the deformation style that characterizes the rifted margin stage. The hanging wall of the last active fault, fc, is sufficiently thin (~16.5 km; Fig. 4c) for most of the crust to behave brittlely[14], promoting strain localization[25]. We propose that a new fault forms at the position favourable for strain localization (f1 in Fig. 4d), cutting the hanging wall to fc, slipping and further thinning the crust until it locks. Subsequent faults form sequentially, following the same pattern (Fig. 4e–k). To model the rifted margin stage, we use geometry, rotation angle and basement subsidence of faults f1–f7 measured on IAM11 (Fig. 2). After f1 stops, f2 forms in the hanging wall of f1. Slip and rotation on f2 deforms its footwall, changing the geometry of the deep segment of the inactive f1 from planar to slightly concave up (Fig. 4e–k), comparable to imaged fault geometry (Fig. 2a, b).

Sequential formation of single faults in the hanging wall of preceding faults, characterizing the rifted margin stage, efficiently thins the crust in comparison with an array of coeval faults, because successive single faults cut crust thinned beforehand by the preceding fault. It also leads to an asymmetric margin structure: one side of the model has inward-dipping faults and gentle Moho shoaling, resembling the WIM, and the conjugate side is characterized by more abrupt Moho shoaling and less faulting (Fig. 4k). To test whether the model reproduces the NFM structure, we include thinning by faulting at the NFM (Fig. 3). Although timing is not well constrained, assuming that faults worked sequentially yields a structure resembling the NFM (Fig. 4k). The model thus reproduces the structure of rifted margins, providing insight into several unresolved problems.

A first model outcome is that it resolves the discrepancy between horizontal extension factors attained by faulting (β_f) and crustal thinning factors[2,3] (β_{cf}). If all faults function simultaneously, β_{cf} at highly
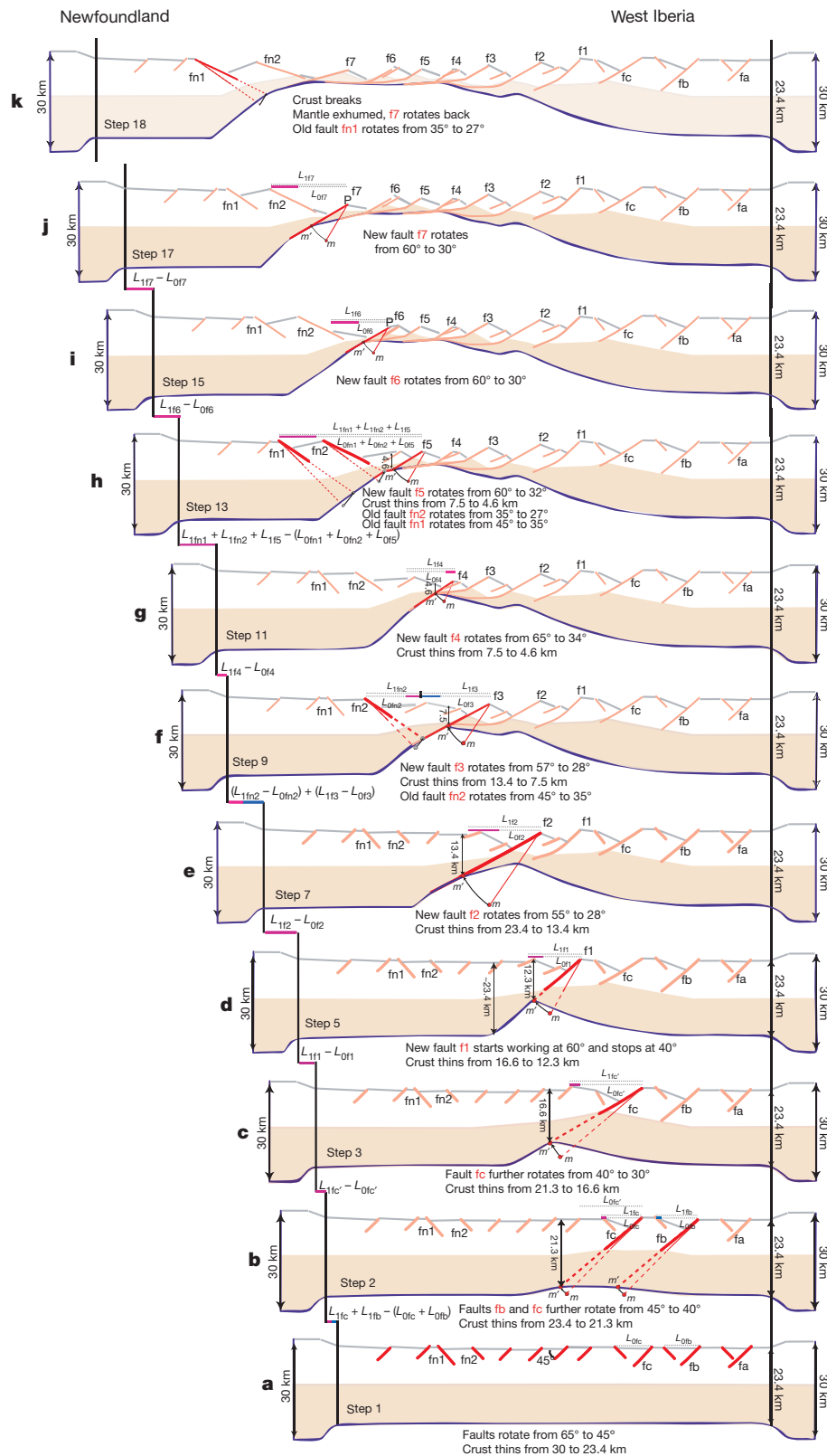
Figure 4 | Selected steps of the balanced cross-section of the kinematic model showing development of crustal thinning and margin asymmetry. Model is fixed at right-hand end. Basement is overlain by pre-rift sediment (in grey). Red and orange lines are active and inactive faults, respectively. Thin red lines mark the initial fault geometry. Black vertical arrows to the left of faults are crustal thicknesses. Blue line marks the Moho. Red dashed lines are projection of faults through ductile crust, used to geometrically estimate Moho movement (Supplementary Fig. 2). Fault activity and rotation angles in **a–c** are based on rift basins[20], and those in **d–k** were measured on IAM11 (Fig. 2). Pale orange shading indicates petrological lower crust, which switches from ductile (**a–c**) to progressively more brittle at the basin centre[14] (**d–k**). At each step, horizontal extension equals extension by active faults (indicated by pink and blue horizontal lines). L_{0fi} and L_{1fi} are the horizontal lengths of a fault block before and after extension of fault fi (i = a, b, c, 1, 2, 3, 4, 5, 6, 7, n1, n2). $L_{0fc'}$ and $L_{1fc'}$ are the horizontal lengths of the hanging-wall block to fault fc before and, respectively, after the fault movement occurring between steps 2 and 3. All reconstruction steps (including some omitted here) are shown in Supplementary Fig. 3. Geometrical rules are described in Supplementary Fig. 2 and Methods.
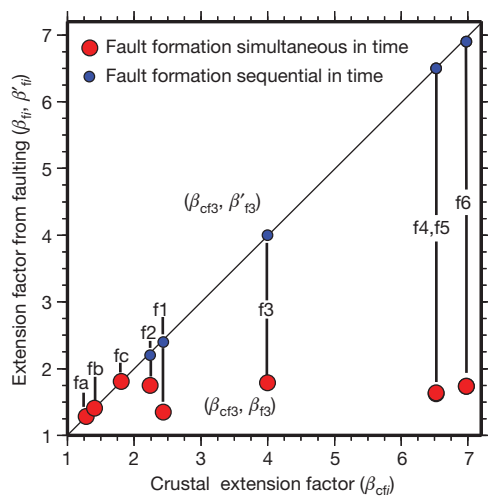
Figure 5 | Extension factors from model in Fig. 4. Crustal thinning factor (β_{cfi}) versus horizontal extension factors from faulting (β_{fi}, red) and combined horizontal extension factor from sequential faulting (β'_{fi}, blue): β_{cfi} is the ratio of initial thickness ($Z_0 = 30$ km) to crustal thickness in the hanging wall to fault fi, $\beta_{fi} = L_{1fi}/L_{0fi}$ (Fig. 4) and $\beta_{cfi} > \beta_{fi}$. If faults function simultaneously, β_{fi} cannot be reconciled with β_{cfi} for faults f1–f6, giving rise to the so-called extension paradox. If faults function sequentially, as in Fig. 4d–k, each new fault cuts through crust that has already been thinned. To account for sequentially, we define the combined extension factor: $\beta'_{fi} = \Pi_{j=n,i-1}\beta_{fj}$. Here $\beta_{fn}, \ldots, \beta_{fi-1}$ are horizontal extension factors of faults that contributed to crustal thinning in the immediate footwall of fi. In our model, crustal thinning is prescribed to equal horizontal extension at each time step (Supplementary Fig. 2), such that sequential faulting results in β'_{fi} equalling β_{cfi}, mapping along the line of no extension discrepancy.

stretched margins (with $\beta_{cf} > 1.8$) is greater than the extension factor achievable with Andersonian faulting (Fig. 5). However, faults forming sequentially when $\beta_{cf} > 1.8$ resolve the paradox, because each fault cuts the hanging wall of the previous fault, adding localized further extension, such that the combined extension factor β'_f is equal to β_{cf} (Fig. 5). Crustal thinning of NFM is not explained by faults imaged there, because it was caused by faults currently located in the WIM (Fig. 4d–k).

A second model outcome concerns the asymmetric structure characteristic of conjugate margins[5,6], which is typically explained by extension along long-lasting, crustal-scale detachment faults[1] active from early in the rifting at mechanically difficult low angles[26]. In our model, asymmetry arises from strain localization promoting one direction of faulting during the rifted margin stage, as imaged on IAM11. We speculate that rheology controls the transition from distributed to localized extension. During early rifting, thick crust with ductile lower crust tends to distribute deformation on small faults over a wide area. Increasing thinning progressively makes the lower crust—and, thus, the entire crust—brittle[14], promoting strain localization in the same way as at slow-spreading ocean ridges, where most extension occurs on a dominant fault cutting towards the thinnest lithosphere at the rift centre. Furthermore, the model predicts exhumation of lower crust to successively shallower levels by faults f1–f7. Evidence in support of this comes from rocks drilled at the edge of the continental crust in the WIM[15].

A related outcome concerns the style of faulting during break-up. In the WIM, a subhorizontal structure imaged on profiles (for example underlying blocks B4–B7 in Fig. 2) has been interpreted as a detachment fault[27]. Similar interpretations in other areas of the WIM and elsewhere consider those structures to be critical during the late rifting[5,8,10,19,28]. Sequential faulting creates similar low-angle structures by passively rotating segments of inactive faults contained within the footwall block of active faults dipping at >28° (Fig. 4 and Methods). Thus, Andersonian faulting remains a viable alternative to widespread detachment faulting controlling break-up.

A last corollary is that the model may resolve two sediment-deposition paradoxes. It has been proposed that syn-rift sediment is surprisingly scarce in the WIM for the time elapsed between pre-rift deposition and initiation of seafloor spreading[18]. Although this controversial interpretation may partially arise from poor imaging (see, for example, Supplementary Fig. 1), sequential faulting implies that the syn-rift sediment on every tilted block is different in age, such that the total syn-rift stratigraphic log is thicker than if all faults work coeval. Also, the model naturally explains the observation that pre-rift or early syn-rift sediments deposited in shallow water, that is, on little-thinned crust, are found overlying extremely thin crust at great water depth[17,24] (Fig. 4).

In contrast to conventional wisdom, a model based on structures unambiguously imaged by seismic methods explains the architecture of conjugate rifted margins formed by extension of cold lithosphere, showing that crustal thinning and asymmetry can both be explained to a first degree by Andersonian faulting.

METHODS SUMMARY

The greater penetration and higher vertical (temporal) and horizontal (spatial) resolution of modern multichannel seismic reflection data, the correct geometry of pre-stack depth-migrated images (Supplementary Fig. 1) and seismic velocity information obtained during depth migration have allowed us to re-evaluate the stratigraphy of fault blocks. The geometry of strata above the S reflector was recently studied[27], and we have extended the analysis to sediment units across the segment of the deep Galicia margin where most drilling occurred, and calibrated the images with Ocean Drilling Program Leg 103 results[17] (Fig. 2c).

Additionally, we have developed a methodology for the kinematic reconstruction of crustal-scale balanced cross-sections that accounts for brittle and initial ductile deformation of the lower crust, subsequent lower-crust embrittlement during rifting[14], progressive focusing of faulting towards the basin centre[20,22] and the geometry of seismically imaged structures. Commercially available balancing software requires a fault detachment level, decoupling upper crustal deformation from lower crustal deformation, and cannot simulate the fault-controlled crustal-scale thinning imaged in our data. We achieve this crustal-scale thinning by moving Moho each extension step to a position where thinning across the active fault equals horizontal extension by the fault (Supplementary Fig. 2). This is done by rotating the point where the fault, or its projection, intersects the Moho by an angle equal to the fault rotation angle (Supplementary Fig. 2). The reconstruction (Supplementary Fig. 3) follows simple geometrical rules, described in Methods and Supplementary Fig. 2, producing a balanced cross-section with area conservation of better than 90%. The final model (Fig. 4k and Supplementary Fig. 3r) resembles the observed asymmetric architecture of conjugate margins (Supplementary Fig. 4). Our model underestimates thinning in NFM, possibly because seismic line SCREECH-1 is not perfectly conjugate to IAM11 (plate reconstruction uncertainties are tens of kilometres), and is somewhat oblique to faults, with the result that imaged fault dip is apparent and extension by faulting is underestimated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lister, G. S., Etheridge, M. A. & Symonds, P. A. Detachment models for the formation of passive continental margins. *Tectonics* **10,** 1038–1064 (1991).
2. Davis, M. & Kusznir, N. J. in *Proc. NSF Rifted Margins Theor. Inst.* (ed. Karner, G. D.) 92–136 (Columbia Univ. Press, 2004).
3. Reston, T. J. Extension discrepancy at North Atlantic nonvolcanic rifted margins: depth-dependent stretching or unrecognized faulting? *Geology* **35,** 367–370 (2007).
4. Sibuet, J. C. New constraints on the formation of the non-volcanic continental Galicia–Flemish Cap conjugate margins. *J. Geol. Soc. Lond.* **149,** 829–840 (1992).
5. Hopper, J. R. *et al.* Continental break-up and the onset of ultraslow seafloor spreading off Flemish Cap on the Newfoundland rifted margin. *Geology* **32,** 93–96 (2004).
6. Louden, K. E. & Chian, D. The deep structure of non-volcanic rifted continental margins. *Phil. Trans. R. Soc. Lond. A* **357,** 767–804 (1999).
7. Driscoll, N. W. & Karner, G. D. Lower crustal extension across the Northern Carnarvon basin, Australia: evidence for an eastward dipping detachment. *J. Geophys. Res.* **103,** 4975–4991 (1998).
8. Lavier, L., & Manatschal, G. A mechanism to thin the continental lithosphere at magma-poor margins. *Nature* **440,** 324–328 (2006).
9. Marrett, R. & Allmendinger, R. Amount of extension on "small" faults: an example from the Viking graben. *Geology* **20,** 47–50 (1992).

10. Whitmarsh, R. B., Manatschal, G. & Minshull, T. A. Evolution of magma-poor continental margins from rifting to seafloor spreading. *Nature* **413**, 150–154 (2001).
11. Ercilla, G. *et al.* Contourites in the Galicia Bank region (NW Iberian Atlantic). *Mar. Geophys. Res.* (in the press).
12. Soliva, R., Benedicto, A. & Maerten, L. Spacing and linkage of confined normal faults: importance of mechanical thickness. *J. Geophys. Res.* **111,** B01402 (2006).
13. Ackermann, R. V., Schlische, R. W. & Withjack, M. O. The geometric and statistical evolution of normal fault systems: an experimental study of the effects of mechanical layer thickness on scaling laws. *J. Struct. Geol.* **23**, 1803–1819 (2001).
14. Pérez-Gussinyé, M. & Reston, T. J. Rheological evolution during extension at passive non-volcanic margins: onset of serpentinization and development of detachments leading to continental break-up. *J. Geophys. Res.* **106**, 3961–3975 (2001).
15. Gardien, V. *et al.* The evolution of amphibolites from Site 1067, ODP Leg 173 (Iberia Abyssal Plain): Jurassic rifting to the Pyrenean compression. *Spec. Publ. Geol. Soc. (Lond.)* **187**, 191–208 (2001).
16. Muentener, O., Hermann, J. & Tromsdorff, V. Cooling history and exhumation of lower-crustal granulite and upper mantle (Malenco, eastern central Alps). *J. Petrol.* **41,** 175–200 (2001).
17. Boillot, G. & Winterer, E. L. Drilling on the Galicia Margin: retrospect and prospect. *Proc. ODP Sci. Results* **103,** 809–828 (1988).
18. Wilson, R. C. L., Manatschal, G. & Wise, S. Rifting along non-volcanic passive margins: stratigraphic and seismic evidence from the Mesozoic of the Alps and Western Iberia. *Spec. Publ. Geol. Soc. (Lond.)* **187**, 429–452, 2001.
19. Péron-Pinvidic, G., Manatschal, G., Minshull, T. A. & Sawyer, D. S. Tectonosedimentary evolution of the deep Iberia-Newfoundland margins: evidence for a complex break-up history. *Tectonics* **26,** TC2011 (2007).
20. Cowie, P. A., Underhill, J. R., Behn, M. D., Lin, J. & Gill, C. E. Spatio-temporal evolution of strain accumulation derived from multi-scale observations of Late Jurassic rifting in the northern North Sea: a critical test of models for lithospheric extension. *Earth Planet. Sci. Lett.* **234,** 401–419 (2005).
21. Walsh, J. J. *et al.* Strain localisation and population changes during fault system growth within the Inner Moray Firth, northern North Sea. *J. Struct. Geol.* **25,** 307–315 (2003).
22. Gawthorpe, R. L. *et al.* Normal fault growth, displacement localisation and the evolution of normal fault populations: the Hamman Faraun fault block, Suez Rift, Egypt. *J. Struct. Geol.* **25,** 1347–1348 (2003).
23. Mattei, M. *et al.* Tectonic evolution of fault-bounded continental blocks: comparison of paleomagnetic and GPS data in the Corinth and Megara basins (Greece). *J. Geophys. Res.* **109,** B02106 (2004).
24. Hopper, J. & Buck, R. Styles of extensional decoupling. *Geology* **26**, 699–702 (1998).
25. Huismans, R. S. & Beaumont, C. Symmetric and asymmetric lithospheric extension: Relative effects of frictional-plastic and viscous strain softening. *J. Geophys. Res.* **108,** 2496 (2003).
26. Sibson, R. H. Fluid involvement in normal faulting. *J. Geodyn.* **29,** 469–499 (2000).
27. Reston, T. J. *et al.* Movement along a low-angle normal fault: the S reflector west of Spain. *Geochem. Geophys. Geosyst.* **8,** Q06002 (2007).
28. Manatschal, G. New models for evolution of magma-poor rifted margins based on a review of data and concepts from West Iberia and the Alps. *Int. J. Earth Sci.* **93,** 432–466 (2004).
29. Srivastava, S. P., Sibuet, J.-C., Cande, S., Roest, W. R. & Reid, I. R. Magnetic evidence for slow spreading during the formation of the Newfoundland and Iberian margins. *Earth Planet. Sci. Lett.* **182,** 61–76 (2000).
30. Whitmarsh, R. B. *et al.* The ocean-continent boundary off the western continental margin of Iberia: crustal structure west of Galicia Bank. *J. Geophys. Res.* **101,** 28291–28314 (1996).

METHODS

Pre-stack depth-migrated seismic reflection records have allowed the re-evaluation of both the fault-block sediment stratigraphy in the segment of the deep Galicia margin where most drilling has occurred[17] and the tectonic structure formed during rifting. On the basis of this reinterpretation, we have developed a methodology for the kinematic reconstruction of crustal-scale balanced cross-sections.

Kinematic reconstruction. A first set of reconstruction rules (rules 1–6) was imposed to define the evolution of deformation in the crustal scale. These rules are based on published, reasonably well-accepted constraints (rules 1–5), and on the geometry of structures imaged on seismic lines IAM11 and SCREECH-1 (rule 6). We describe these rules and how they define how extension progresses to final continental break-up.

(1) At the start of rifting, the crust is ∼30 km thick and has a temperature of ∼500 °C at the Moho, which is appropriate for rifting conditions at the Galicia margin[8,14,16]. Numerical modelling following ref. 14 shows that given those initial conditions and an appropriate extension velocity for the west Iberia/Newfoundland system, the lower crust (orange shading in Fig. 4, and Supplementary Figs 2–4) progressively becomes brittle, and becomes entirely brittle when the total crust is ∼12–14 km thick (Supplementary Fig. 3h–i). (2) During early rifting phases, at the basin stage (Fig. 4a–c and Supplementary Fig. 3a–c), strain in faults gradually focuses into the basin centre[20,22]. (3) During this basin stage, faults start at 65° and rotate to 45°–30° following Andersonian principles[31,32] until all faults lock up (Fig. 4a–c and Supplementary Fig. 3a–c). (4) We assume that the following margin stage (Fig. 4d–k and Supplementary Fig. 3d–r) starts with the creation of a new fault, after all previous faults have locked up, located where the crust is thinnest (Supplementary Fig. 3d). We choose that location because deformation is expected to focus on the segment with the smallest elastic thickness. (5) During the progression of extension, as the lower crust switches from ductile to brittle deformation, it is first partially and later entirely cut by faults (faults f1–f7 in Fig. 4d–k and Supplementary Fig. 3d–r), and is exhumed at the thinnest fault blocks (faults f5–f7 in Fig. 4h–k and Supplementary Fig. 3m–r) as observed at drilled continental margins and exhumed margin analogues[28]. (6) To create the margin stage geometry, fault spacing, initial and final fault rotation angles and basement subsidence are all directly taken from measured structures imaged at profile IAM11 (Fig. 2). In this manner, isostasy of large-offset faults is embedded in the reconstruction.

The second set of reconstruction rules (rules 7–15) is imposed to maintain common geometric rules in all steps of the model, and conforms with the deformation principles defined by the first set of rules.

(7) Large active faults are planar in the brittle domain[33,34]. (8) During model construction, we choose horizontal extension to be accomplished by rotation of the active faults. Each fault rotates with respect to an axis at its top end. It can be proved that the resulting geometry of the crust and fault is independent of the location chosen for the rotation axis along the fault. We choose the fault top as the rotation axis to simplify the reconstruction. (9) The horizontal extension of the model at a time step is equal to the cumulative horizontal extension of all faults active at that time step. (10) A fault has the same length along the hanging wall and the footwall. (11) During faulting, the top and base of brittle hanging-wall crust are offset by faulting, but when the lower crust is ductile, fault offset attenuates across the ∼3-km-thick brittle ductile transition (BDT; Supplementary Fig. 2). (12) The lower crust undergoes deformation beneath the segments where the upper crust extends, such that lower crustal deformation changes from a broad region (Fig. 4a, b and Supplementary Fig. 3a, b) to a narrow region in parallel with the change in strain distribution from multiple faults to one fault as extension progresses (Fig. 4c–k and Supplementary Fig. 3c–r). (13) Crustal thinning by an active fault equals horizontal extension by the fault. This is attained by rotating the location where (the projection of) the fault intersects the Moho by an angle equal to the rotation angle of the fault (Supplementary Fig. 2). (14) We assume that block deformation decreases away from faults. Footwall deformation is approximated by an affine transformation involving shear and stretch of triangle POm into POm′ (Supplementary Fig. 2), which is a reasonable geometrical approximation to real footwall deformation[35]. The distance over which the footwall deformation associated with fault slip decreases to zero possibly depends on the elastic thickness and, thus, on the evolving rheology. However, there is no established geometric principle to determine how far footwall deformation extends, and we considered two modes depending on crustal rheology. When the lower crust is ductile, footwall deformation occurs underneath the fault plane and the area increase from POm to POm′ is accomplished by ductile deformation (Supplementary Figs 2a–c and 3a–d). When the entire crust is brittle (Supplementary Figs 2d, e and 3f, r), footwall deformation affects the active fault footwall block and extends to older fault blocks located underneath the active fault. Here the area increase from POm

to POm′ is <3%, a result easily explained by subseismic internal block deformation[9]. Therefore, changes in fault geometry from planar to listric occurring in the thinner crust of the margin stage result from block deformation (see, for example, Supplementary Figs 2d, e and 3f–r). (15) The ductile parts of the lower crust in the hanging wall deform to account for extension with area preservation (see Supplementary Fig. 2 for details) in a manner similar to the cantilever model[36]; but in that model all faults rotate simultaneously, so it does not explain the extension discrepancy at high stretching factors.

The reconstruction reproduces the asymmetric geometry imaged on IAM11 and SCREECH-1 (Supplementary Fig. 4), and the regional asymmetry observed at other non-volcanic (cold-lithosphere) conjugate margins surveyed using coincident deep-penetrating wide-angle seismic data and multichannel seismic reflection images, such as the Porcupine basin rift flanks[37], south Labrador Sea/Greenland[38,39], Armorica/Flemish Cap[6,38], the Galicia Interior basin flanks[40] and Nova Scotia/Morocco[41], and even magmatically robust systems such as the Gulf of Aden[42] and Woodlark basin[43], are clearly asymmetrical.

Fault and coeval Moho rotation (at, for example, f5; Supplementary Fig. 3m) may cause rotation of the deep segment of a neighbouring inactive fault (for example f4 in Supplementary Fig. 3l). In thick crust, large faults tend to be distant, so fault rotation does not significantly affect the planar geometry of older inactive ones (fa, fb and fc; Fig. 4a–c and Supplementary Fig. 3a–c). As the crust thins and faults form at smaller spacing, rotation of an active fault modifies the geometry of the plane of a neighbouring inactive fault in its deep segment. The inactive fault plane is passively rotated to a smaller angle, producing an overall listric fault geometry, or even a subhorizontal appearance (Fig. 4e–k and Supplementary Fig. 3g–r). In particular, our model reproduces the break-away and subhorizontal geometry of a previously interpreted detachment fault. But our model does not require a mechanically controversial very-low-angle fault activity (the S reflector, imaged under blocks B4–B7 in Fig. 2, corresponding to blocks bounded by faults f4–f6 in Fig 3 and Supplementary Fig. 3).

The kinematics of fault blocks B4–B7 might resemble the 'roll and hinge' model[44]. However, our model is different because it has no hinge fixed in space from which successive slivers of crust are sliced. Additionally, the model reproduces—using one common Andersonian mechanistic framework—the change in fault geometry that has been related to increasing crustal extension at margins: the model evolves from deformation at distributed planar faults in little-extended sectors to deformation on faults that, owing to passive rotation, rest at low angles at highly extended sectors.

31. Anderson, E. M. The dynamics of faulting. *Trans. Edinb. Geol. Soc.* **8,** 387–402 (1905).
32. Célérier, B. Seeking Anderson's faulting in seismicity: a centennial celebration. *Rev. Geophys.* **46,** RG4001 (2008).
33. Thatcher, W. & Hill, D. P. A simple model for the fault-generated morphology of slow-spreading mid-ocean ridges. *J. Geophys. Res.* **100,** 561–570 (1995).
34. Jackson, J. A. Active normal faulting and crustal extension. *Spec. Publ. Geol. Soc (Lond.)* **28,** 3–17 (1987).
35. Westaway, R. & Kusznir, N. Fault and bed 'rotation' during continental extension: block rotation or vertical shear? *J. Struct. Geol.* **15,** 753–770 (1993).
36. Kusznir, N. J. Marsden, G. & Egan, S. S. A flexural-cantilever simple-shear/pure-shear model of continental lithosphere extension: applications to the Jeanne d'Arc Basin, Grand Banks and Viking Graben, North Sea. *Spec. Publ. Geological Society (Lond.)* **56,** 41–60 (1991).
37. Reston, T. J. *et al.* Extreme crustal thinning in the south Porcupine Basin and the nature of the Porcupine Median High: implications for the formation of non-volcanic rifted margins. *J. Geol. Soc. Lond.* **161,** 783–798 (2004).
38. Keen, C., Peddy, C., de Voogd, B. & Matthews, D. Conjugate margins of Canada and Europe: results from deep reflection profiling. *Geology* **17,** 173–176 (1989).
39. Chian, D., Keen, C., Reid, I. & Louden, K. E. Evolution of nonvolcanic rifted margins: new results from the conjugate margins of the Labrador Sea. *Geology* **23,** 589–592 (1995).
40. Pérez-Gussinyé, M., Ranero, C. R., Reston, T. J. & Sawyer, D. Mechanisms of extension at nonvolcanic margins: evidence from the Galicia interior basin, west of Iberia. *J. Geophys. Res.* **108,** 2245 (2003).
41. Maillard, A., Malod, J., Thiebot, E., Klingelhoefer, F. & Rehault, J.-P. Imaging a lithospheric detachment at the continent–ocean crustal transition off Morocco. *Earth Planet. Sci. Lett.* **241,** 686–698 (2006).
42. Fournier, M., Huchon, P., Khanbari, K. & Leroy, S. Segmentation and along-strike asymmetry of the passive margin in Socotra, eastern Gulf of Aden: are they controlled by detachment faults? *Geochem. Geophys. Geosyst.* **8,** Q03007 (2007).
43. Kington, J. D. & Goodliffe, A. M. Plate motions and continental extension at the rifting to spreading transition in Woodlark Basin, Papua New Guinea: can oceanic plate kinematics be extended into continental rifts? *Tectonophysics* **458,** 82–95 (2008).
44. Buck, W. R. Flexural rotation of normal faults. *Tectonics* **7,** 959–973 (1988).

LETTER

ON and OFF pathways in *Drosophila* motion vision

Maximilian Joesch[1]*, Bettina Schnell[1]*, Shamprasad Varija Raghu[1], Dierk F. Reiff[1] & Alexander Borst[1]

Motion vision is a major function of all visual systems, yet the underlying neural mechanisms and circuits are still elusive. In the lamina, the first optic neuropile of *Drosophila melanogaster*, photoreceptor signals split into five parallel pathways, L1–L5[1]. Here we examine how these pathways contribute to visual motion detection by combining genetic block and reconstitution of neural activity in different lamina cell types with whole-cell recordings from downstream motion-sensitive neurons[2,3]. We find reduced responses to moving gratings if L1 or L2 is blocked; however, reconstitution of photoreceptor input to only L1 or L2 results in wild-type responses. Thus, the first experiment indicates the necessity of both pathways, whereas the second indicates sufficiency of each single pathway. This contradiction can be explained by electrical coupling between L1 and L2, allowing for activation of both pathways even when only one of them receives photoreceptor input. A fundamental difference between the L1 pathway and the L2 pathway is uncovered when blocking L1 or L2 output while presenting moving edges of positive (ON) or negative (OFF) contrast polarity: blocking L1 eliminates the response to moving ON edges, whereas blocking L2 eliminates the response to moving OFF edges. Thus, similar to the segregation of photoreceptor signals in ON and OFF bipolar cell pathways in the vertebrate retina[4], photoreceptor signals segregate into ON-L1 and OFF-L2 channels in the lamina of *Drosophila*.

Neurons responding to visual motion in a directionally selective way are found in a vast number of animals and brain regions, ranging from the retina of rabbits[5] to the visual cortex of macaques[6]. In flies, large-field motion-sensitive neurons are located in the third neuropile layer, the lobula plate (Fig. 1a), and are thought to be involved in visual flight control[7]. These lobula plate tangential cells are preferentially sensitive to vertical (VS cells) and horizontal (HS cells) motion, respectively. They depolarize when stimulated by motion along their preferred direction (PD motion) and hyperpolarize during motion along the opposite, null direction (ND motion). In the first neuropile, the lamina, photoreceptors R1–R6 provide input, directly or indirectly, onto five different monopolar cells (L1–L5)[1,8] using histamine as their transmitter[9]. L1–L5 send their axons into the medulla where neurons compute the direction of motion in accordance with the Reichardt model[10]. Such motion detectors then provide excitatory and inhibitory input onto the dendrites of lobula plate tangential cells[2,3,7]. However, the neural circuitry presynaptic to the tangential cells represented by the Reichardt detectors has so far escaped a detailed analysis, because of the small size of the columnar neurons. We set out to elucidate the cellular implementation of the Reichardt model of visual motion detection starting from the lamina, asking which of the various neurons provide input to the motion detection circuitry. Previous studies addressing this question in *Drosophila* used behavioural read-outs to test for effects of blocking and rescuing of specific lamina cells[11–13]. To get closer to the circuit in question, we used the Gal4[14] or Split-Gal4[15]/UAS system (Fig. 1b) and combined genetic intervention in different lamina neurons with electrophysiological recordings from lobula plate tangential cells.

We recorded from HS and VS cells and blocked the output of lamina neurons L1 and L2 by targeted expression of *shibire*[ts] (ref. 16). Control flies (black traces in Fig. 2a–d) always revealed strong and reliable
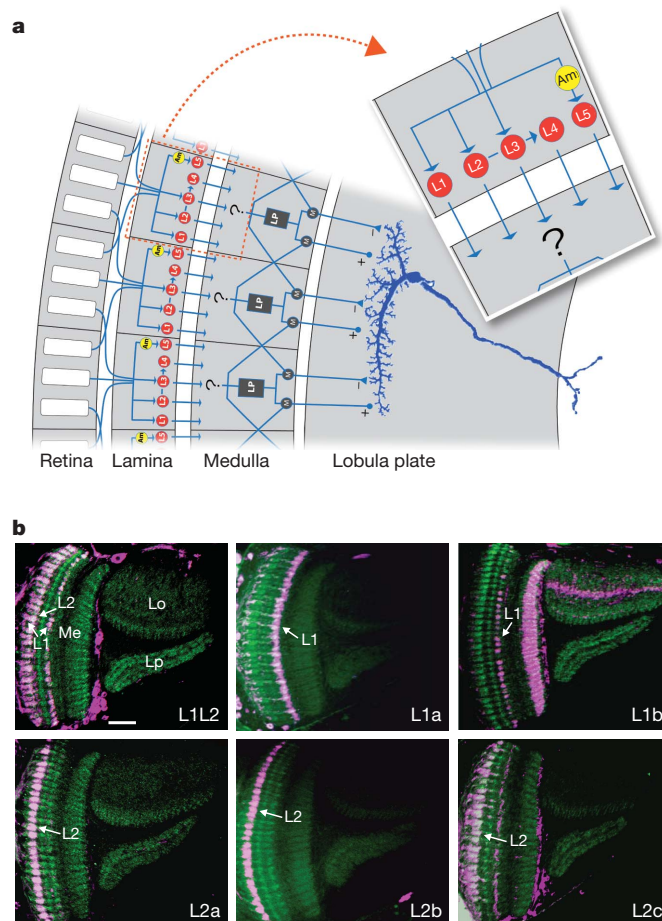
Figure 1 | Fly optic lobe and *GAL4* driver lines. a, Schematic of the fly optic lobe (frontal view). By means of neural superposition, photoreceptors R1–R6 (only three are shown) connect to five different lamina cells. L1–L3 and the amacrine cell (Am) receive direct input from the photoreceptors, whereas L4 and L5 receive indirect input via L2 and the amacrine cell, respectively (simplified after data in ref. 1). Within the medulla, a circuit specified only in algorithmic form (Reichardt detector[10]) transforms signals from adjacent sampling points into directionally selective output signals. Each such detector consists of two mirror-symmetrical subunits, the output signals of which provide excitatory and inhibitory input to lobula plate tangential cells, respectively. In each subunit, the signal derived from one sampling point is low-pass filtered (LP) and subsequently multiplied (M) with the instantaneous signal derived from the neighbouring point. Which of the lamina cells feed into the motion detection circuit and what signals they provide is the central question of this study. **b**, Expression pattern in the optic lobes of the *GAL4* and *Split-GAL4* driver lines used in this study. Cells targeted by the driver line were visualized by expression of membrane-tagged DsRed (magenta). The neuropile was labelled using antisera against Dlg, a postsynaptic marker protein (green). Besides the lamina neurons indicated, all lines label, although less intensely, additional cells. For driver lines L1a, L2a and L2c, see ref. 18. For driver line L1b we found co-expression in some Pm, Mt, Lt and Tlp cells; for driver line L2b we found co-expression in Tm4 cells; and for driver line L1L2 we found co-expression in Tm5, Tlp and T3 cells. (See ref. 30 for description of cell types.) The genotype of each driver line is given in Methods. Scale bar, 20 μm. Horizontal optical sections: Lo, lobula; Lp, lobula plate; Me, medulla.

[1]MPI for Neurobiology, Department of Systems and Computational Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany.
*These authors contributed equally to this work.
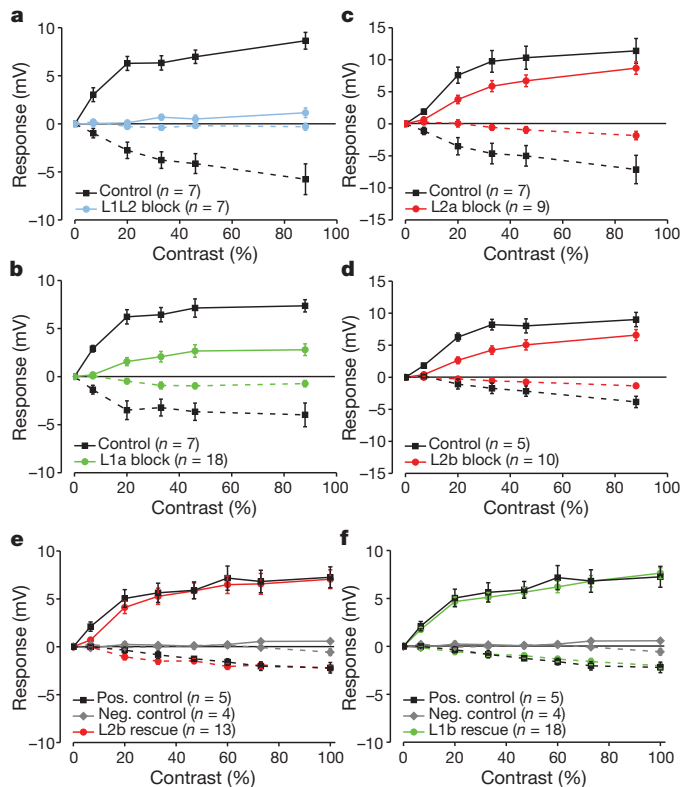
Figure 2 | Tangential cell responses to moving large field gratings of various contrasts. a–d, Lamina neurons L1 and L2 (**a**), only L1 (**b**) and only L2 (**c**, **d**) were blocked by cell-specific expression of *shibire*[ts]. Control flies (in black) had the same genotype as experimental flies, but were kept at permissive temperature throughout, whereas experimental flies (in colour) were put to restrictive temperature 1 h before the experiment. **e**, **f**, L1 (**e**) and L2 (**f**) pathways were rescued by cell-specific expression of the wild-type histamine receptor (*ort*) in an *ort*-null mutant. Data from positive control flies (heterozygous wild-type background) are shown in black; data from negative control flies (*ort*-null mutants) in grey; data from rescue flies in colour. Data from positive and negative control flies are identical in **e** and **f**. Responses to grating motion along the preferred direction are shown as solid lines; responses to motion along the null direction are shown as dashed lines. Data represent the mean ± s.e.m. obtained from *n* animals (*n* is indicated in each panel). Because recordings from HS and VS cells revealed no difference, both groups were pooled. Blocking L1 or L2 significantly reduces the motion response (**b–d**), thus each pathway seems to be necessary. However, rescuing either one of them fully restores the wild-type response (**e**, **f**). Thus, each pathway appears sufficient.

responses to a moving grating, saturating for increasing contrast levels for both PD motion (solid lines) as well as ND motion (dashed lines). Blocking both L1 and L2 led to a complete loss of motion responses even at the highest pattern contrast (Fig. 2a, driver line L1L2, blue traces). Blocking only L1 strongly reduced PD and ND responses for all contrasts tested (Fig. 2b, driver line L1a, green traces). Blocking L2 using two different driver lines moderately reduced the responses at all contrast levels (L2a, Fig. 2c; L2b, Fig. 2d, red traces). To test whether the temperature shift alone could lead to altered motion responses, flies that had the same genotype as experimental flies except for the *GAL4* driver gene were put to restrictive temperature 1 h before the experiment. The responses of these flies were indistinguishable from the ones of the other control flies (*n* = 5, data not shown). Together, these results indicate that L1 and L2 are necessary for wild-type responses to grating motion.

In a complementary approach, we selectively rescued photoreceptor input to lamina cells L1 and L2 via targeted expression of the wild-type histamine receptor, encoded by the *ort* gene[17], in an *ort*-null mutant background[11,18]. Given the above results from the blocking experiments, rescuing either L1 or L2 pathway should lead to only small

motion responses at best. However, rescuing L2 led to wild-type motion responses at all contrasts tested, for PD motion as well as for ND motion (Fig. 2e; driver line L2b, red traces). The same was true when lamina cells L1 were rescued: again, motion responses were nearly indistinguishable from the ones of 'positive control' flies (Fig. 2f; driver line L1b, green traces). In these positive control flies, no L1- or L2-*GAL4*, but one wild-type *ort*-allele, was present, leading to wild-type motion responses as expected (Fig. 2e, f, black traces). In 'negative control' flies, where either no L1-*GAL4* and L2-*GAL4* or no UAS-*ort* was present in an *ort*-null mutant background, motion responses were literally zero (Fig. 2e, f, grey traces). Thus, blocking L1 or L2 revealed that the output of both L1 and L2 is necessary for wild-type motion responses. Rescuing the pathway of either L1 or L2 indicates, however, that either L1 or L2 is sufficient for a wild-type motion response. This contradiction deserves further investigation.

The blocking and rescuing experiments presented above differ in one important aspect: in one case, the synaptic output of L1 and L2 was blocked, in the other case, the synaptic input to the same cells was rescued. If L1 and L2 receive their input in parallel without any further interactions, both procedures should yield complementary results, which we did not find. Thus, we examined the existence of electrical connections between L1 and L2 by immunolabelling of the innexin protein Shaking B, a member of the gap-junction-forming protein family in flies[19,20]. We found strong immunolabelling within the entire optic lobe including the lamina (Fig. 3a). Furthermore, the basal laminar processes of L1 and L2 appeared to co-localize with the Shaking B immunolabelling (Fig. 3a, insert). Because some dipteran gap junctions were demonstrated to be permeable for neurobiotin[2,21], we injected L1 cells with neurobiotin and looked for co-staining in L2. When a single L1 cell was injected, a clear staining became visible in the adjacent L2 cell as well, identified by its characteristic terminal in medulla layer 2 (Fig. 3b). Injecting L2 led to co-staining of the adjacent L1 cell, identified by its characteristic terminals in medulla layers 1 and 5 (Fig. 3c). We therefore propose that L1 and L2 are electrically coupled via gap junctions.

Gap-junctional coupling between L1 and L2 could, in principle, explain the contradictory results obtained in blocking and rescuing experiments: through electrical coupling, rescuing the photoreceptor input to L1 restores the L2 pathway as well, and vice versa. This explanation, however, requires that the coupling between L1 and L2 provides a sufficiently large input to the respective partner cell. To investigate the strength of the coupling, we expressed an inwardly rectifying potassium channel (Kir2.1)[22] in one of the two lamina cells. When we expressed the potassium channel in L1 alone, motion responses were completely abolished (Fig. 3d, e, driver line L1a, green traces), comparable to the situation when L1 and L2 were blocked by *shibire*[ts] (Fig. 2a, blue traces). A similar finding was obtained when the potassium channel was expressed in L2 cells (Fig. 3d, e, driver line L2c, red traces). These results indicate a strong electrical coupling between L1 and L2 and, thus, resolve the apparent discrepancy between blocking and rescuing experiments.

So far, our data support the view that both L1 and L2 feed, with a somewhat different contribution, into the motion detection circuitry. However, no evidence is provided as to any functional specialization of each of the pathways. As one possibility, lamina cells L1 and L2 might be specifically involved in the analysis of either ON or OFF input signals, in analogy to the vertebrate retina[4]. Because a grating stimulus is composed of many simultaneously moving dark-to-bright (ON edge) and bright-to-dark transitions (OFF edge), this would have escaped our analysis presented above. To investigate this possibility, we presented moving edges of a single polarity to flies in which we blocked the output of lamina cells L1 and L2 by *shibire*[ts]. In control flies, moving ON and OFF edges elicited strong and reliable voltage responses in lobula plate tangential cells during PD and ND motion (Fig. 4a, b, black traces). When the output from L1 was blocked, the response to moving ON edges was literally zero whereas the response to moving OFF edges was
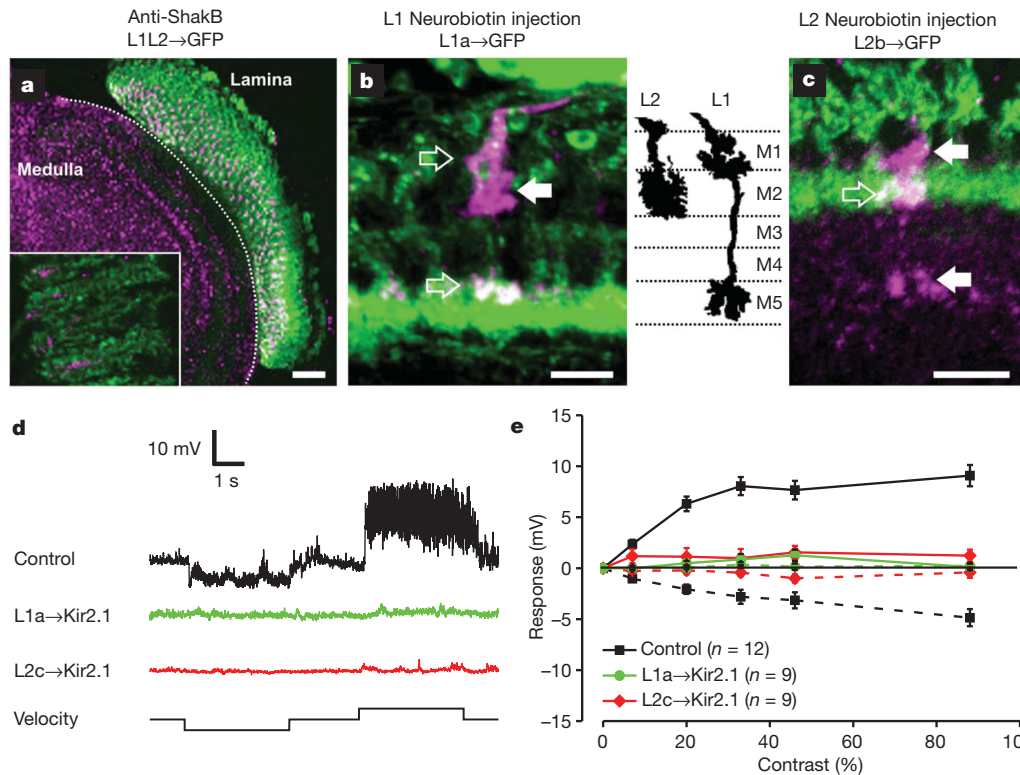
Figure 3 | Lamina cells L1 and L2 are electrically coupled.
a, Immunostaining of the lamina, using the Shaking B antibody (magenta), in an L1L2 line expressing GFP (green). The inset represents a threefold magnification of the large figure showing immunostaining primarily in the proximal part of adjacent lamina cartridges. Scale bar, 25 μm. **b**, Dye-coupling between L1 and L2 cells. An L1 cell was injected with neurobiotin in an L1a line expressing GFP (green). The terminals of a single L2 cell (magenta, white arrow) are visible in the medulla. The open arrows point towards the terminals of the injected L1 cell (white). Similar results were obtained in $n = 6$ flies. **c**, An L2 cell was injected with neurobiotin in an L2b line expressing GFP (green). The terminals of a single L1 cell (magenta, white arrows) are visible in the medulla. The open arrow points towards the injected L2 cell (white). Similar results were obtained in $n = 7$ flies. Scale bars in **b** and **c**, 5 μm. The inset between **b** and

c shows the terminals of Golgi-impregnated L1 and L2 cells (from ref. 30). **d**, **e**, Voltage responses of lobula plate tangential cells in flies expressing an inwardly rectifying potassium channel (Kir2.1) in L1 cells via the driver line L1a, or in L2 cells via the line L2c. For each fly, a single response trace is shown in **d**; the average response is shown in **e**. In **e**, data represent the mean ± s.e.m. of the data, obtained from n flies (n indicated in each panel). Data from control flies are shown in black; data from experimental flies are shown in colour. Responses to grating motion along the preferred direction are shown as solid lines; to motion along the null direction as dashed lines. Because recordings from HS and VS cells revealed no difference, both groups were pooled. Motion responses are almost completely abolished, no matter whether Kir2.1 was expressed in L1 or in L2.

still about 50% of the wild-type response (Fig. 4a, b, top row; driver line L1a, green traces). The opposite was true when the output from L2 was blocked by expressing *shibire*[ts] using two different *GAL4* driver lines: then, the response to moving ON edges was only mildly reduced whereas the response to moving OFF edges was nearly abolished (Fig. 4a, b, second and third row; driver lines L2a and L2b, red traces).

In a pioneering study, and consistent with our results, it was found that rescuing either the L1 or the L2 pathway led to wild-type opto-motor responses at high pattern contrasts[11]. For low contrasts (5–10%), a functional specialization of the L1 and L2 pathway for back-to-front and front-to-back motion was suggested[11], which, however, does not match our data on tangential cell responses in that contrast range (Fig. 2e, f). The first evidence for a role of the L2 pathway in transmitting light OFF signals was obtained in a study on freely walking flies, where blocking L2 impaired turning tendencies in response to contrast decrements[12]. However, our finding that photoreceptor signals in the fly segregate into ON and OFF pathways via L1 and L2 neurons is surprising in so far as, different from ON and OFF bipolar cells of the vertebrate retina[4], both lamina cell types posses the same transmitter receptor[9] and produce similar light responses in their dendrite[23]. This similarity is likely to be increased even further by the gap-junctional coupling between dendritic compartments of L1 and L2, which might help to average out uncorrelated noise both cells receive from photo-receptor R1–R6 input. Subsequently, however, these signals must become differentially rectified. For L2, this has been recently shown

to occur already within the cell, as L2 axon terminals reveal pro-nounced calcium signals selectively in response to light OFF stimuli[24]. Whether this also holds true for L1, or whether the selective respon-siveness of the L1 pathway to light ON stimuli is only acquired further downstream in its postsynaptic neurons[25], is currently not known. On the basis of the co-stratification of columnar neurons[26] as well as 2-deoxyglucose activity labelling[27], L1 and L2 have long been proposed to represent the entry points to two parallel motion pathways in the fly visual system, with L1 synapsing onto medulla intrinsic neuron Mi1 which in turn contacts T4 cells, L2 synapsing onto transmedullar neuron Tm1 which in turn contacts T5 cells, and with T4 and T5 cells finally converging on the dendrites of the lobula plate tangential cells. Our results provide evidence that these two pathways deal specifically with the processing of ON and OFF stimuli. Moreover, splitting a positively and negatively going signal into separate ON and OFF channels alleviates the neural implementation of a multiplication, as postulated by the Reichardt detector. Whereas otherwise, the output signal of the multiplier had to increase in a supra-linear way when both inputs increase as well as when they decrease, dealing with positive signals only in separate multipliers seems to be less demanding with respect to the underlying biophysical mechanism[28]. Whatever this mechanism will turn out to be, our finding about the splitting of the photoreceptor signal into ON and OFF pathways adds to the already described commonalities between the invertebrate and the vertebrate visual sys-tem[29]. Obviously, the selection pressure for an energy-efficient way of
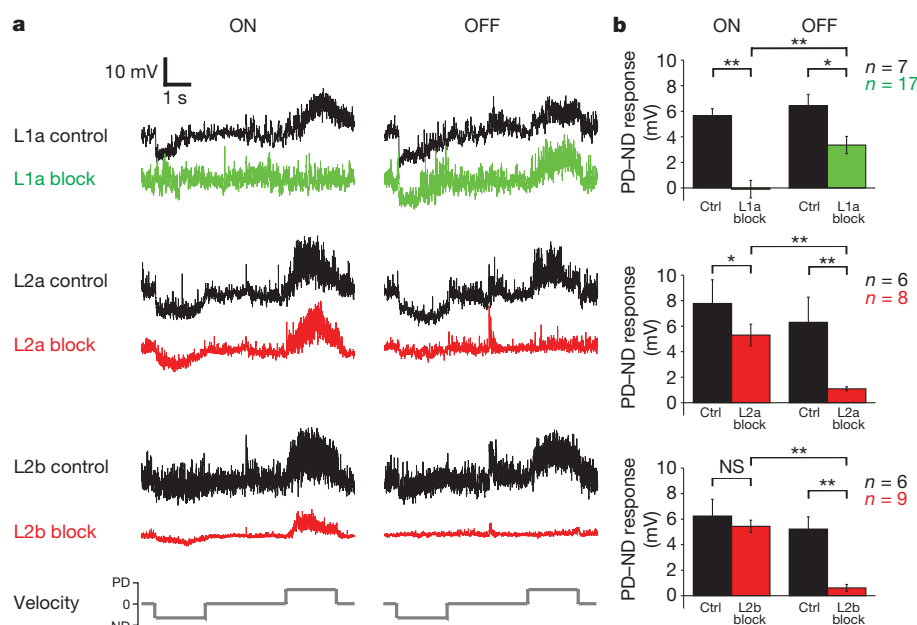
Figure 4 | Voltage responses of lobula plate tangential cells to moving edges of a single polarity (single example response in a; average in b). a, Lamina neurons L1 (top panel) or L2 (middle and bottom panels) were blocked by cell-specific expression of *shibire*[ts]. Control flies (in black) had the same genotype as experimental flies, but were kept and recorded at permissive temperature. Experimental flies (L1a in green; L2a and L2b in red) were shifted to restrictive temperature 1 h before the experiment. **b,** Data represent the mean response (PD − ND) ± s.e.m. of the data obtained from *n* flies (*n* indicated in each panel). Asterisks indicate the significance level of the difference between the mean values: *$P < 0.05$, **$P < 0.001$. NS, not significant. Because recordings from HS and VS cells revealed no difference, both groups were pooled. Blocking L1 abolishes the response to moving ON edges completely, whereas blocking L2 mainly affects the response to moving OFF edges.

encoding light increments and decrements led to rather similar implementations across distant phyla.

METHODS SUMMARY

For the *shibire* experiments, control and experimental flies had identical genotypes and were raised at 18 °C (permissive temperature). Experimental flies were shifted for 1 h to 37 °C (restrictive temperature) directly before the experiment and recorded at room temperature (21 °C). No recovery of the block was detected within the time of recording (see Supplementary Information). For the Kir2.1 experiments, experimental and control flies were raised at 30 °C to boost the expression level of *Kir2.1*. The genotypes of the driver lines are listed in Methods. For whole-cell recording, flies were dissected according to ref. 2. A Ringer's-solution-filled electrode (tip ~4 μm) was used to remove the extracellular matrix and to expose the somata of lobula plate tangential cells. Somata were recognized either by expression of a fluorescent marker or by their location next to a prominent tracheal branch. Recordings in rescue experiments were performed with fluorescently targeted neurons as in ref. 3; recordings in blocking experiments were established under high-contrast optics. For further details of the set-up, see ref. 2. For neurobiotin dye fills, flies expressing mCD8-GFP in L1 or L2 (L1a or L2b) were decapitated. The heads were fixed and lamina cells were injected with a 5 mM Alexa Fluor 568 (A10442; Invitrogen) and 2% neurobiotin (Vector Labs) solution as described in ref. 3. Access to the lamina monopolar cell somata was possible after removal of the fly's retina. A single lamina cell was injected per brain. Neurobiotin was detected by coupling to Streptavidin–Alexa Fluor 568 conjugate (1:100, Invitrogen)[2]. For visual stimulation, we used a custom-built LED arena that covered ~170° (1.4° resolution) of the horizontal and ~100° (1.8° resolution) of the vertical visual field of the fly, allowing refresh rates of up to 600 Hz with 16 intensity levels.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Meinertzhagen, I. A. & O'Neil, S. D. Synaptic organization of columnar elements in the lamina of the wild type in *Drosophila melanogaster. J. Comp. Neurol.* **305,** 232–263 (1991).
2. Joesch, M., Plett, J., Borst, A. & Reiff, D. F. Response properties of motion-sensitive visual interneurons in the lobula plate of *Drosophila melanogaster. Curr. Biol.* **18,** 368–374 (2008).
3. Schnell, B. *et al.* Processing of horizontal optic flow in three visual interneurons of the *Drosophila* brain. *J. Neurophysiol.* **103,** 1646–1657 (2010).
4. Wässle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* **5,** 747–757 (2004).
5. Barlow, H. B., Hill, R. M. & Levick, W. R. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol. (Lond.)* **173,** 377–407 (1964).
6. Dubner, R. & Zeki, S. M. Response properties and receptive fields of cells in an anatomically defined region of superior temporal sulcus in monkey. *Brain Res.* **35,** 528–532 (1971).
7. Borst, A., Haag, J. & Reiff, D. F. Fly motion vision. *Annu. Rev. Neurosci.* **33,** 49–70 (2010).
8. Strausfeld, N. J. *Atlas of an Insect Brain* (Springer, 1976).
9. Hardie, R. C. A histamine-activated chloride channel involved in neurotransmission at a photoreceptor synapse. *Nature* **339,** 704–706 (1989).
10. Reichardt, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* (ed. Rosenblith, W. A.) 303–317 (John Wiley & Sons, 1961).
11. Rister, J. *et al.* Dissection of the peripheral motion channel in the visual system of *Drosophila melanogaster. Neuron* **56,** 155–170 (2007).
12. Katsov, A. Y. & Clandinin, T. R. Motion processing streams in *Drosophila* are behaviorally specialized. *Neuron* **59,** 322–335 (2008).
13. Zhu, Y., Nern, A., Zipursky, S. L. & Frye, M. A. Peripheral visual circuits functionally segregate motion and phototaxis behaviors in the fly. *Curr. Biol.* **19,** 613–619 (2009).
14. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118,** 401–415 (1993).
15. Luan, H., Peabody, N. C., Vinson, C. R. & White, B. H. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron* **52,** 425–436 (2006).
16. Kitamoto, T. Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive *shibire* allele in defined neurons. *J. Neurobiol.* **47,** 81–92 (2001).
17. Gengs, C. *et al.* The target of *Drosophila* photoreceptor synaptic transmission is a histamine-gated chloride channel encoded by *ort* (*hclA*). *J. Biol. Chem.* **277,** 42113–42120 (2002).
18. Gao, S. *et al.* The neural substrate of spectral preference in *Drosophila. Neuron* **60,** 328–342 (2008).
19. Phelan, P., Bacon, J. P., Davies, J. A., Stebbings, L. A. & Todman, M. G. Innexins: a family of invertebrate gap-junction proteins. *Trends Genet.* **14,** 348–349 (1998).

20. Blagburn, J. M., Alexopoulos, H., Davies, J. A. & Bacon, J. P. Null mutation in shaking-B eliminates electrical, but not chemical, synapses in the *Drosophila* giant fiber system: a structural study. *J. Comp. Neurol.* **404,** 449–458 (1999).

21. Haag, J. & Borst, A. Dye-coupling visualizes networks of large-field motion-sensitive neurons in the fly. *J. Comp. Physiol. A* **191,** 445–454 (2005).

22. Johns, D. C., Marx, R., Mains, R. E., O'Rourke, B. & Marban, E. Inducible genetic suppression of neuronal excitability. *J. Neurosci.* **19,** 1691–1697 (1999).

23. Laughlin, S. B. & Osorio, D. Mechanism for neural signal enhancement in the blowfly compound eye. *J. Exp. Biol.* **144,** 113–146 (1989).

24. Reiff, D. F., Plett, J., Mank, M., Griesbeck, O. & Borst, A. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of *Drosophila. Nature Neurosci.* **13,** 973–978 (2010).

25. Takemura, S. Y., Lu, Z. & Meinertzhagen, I. A. Synaptic circuits of the *Drosophila* optic lobe: the input terminals to the medulla. *J. Comp. Neurol.* **509,** 493–513 (2008).

26. Bausenwein, B., Dittrich, A. P. M. & Fischbach, K.-F. The optic lobe of *Drosophila melanogaster*. II. Sorting of retinotopic pathways in the medulla. *Cell Tissue Res.* **267,** 17–28 (1992).

27. Bausenwein, B. & Fischbach, K.-F. Activity labeling patterns in the medulla of *Drosophila melanogaster* caused by motion stimuli. *Cell Tissue Res.* **270,** 25–35 (1992).

28. Hassenstein, B. & Reichardt, W. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Ruesselkaefers Chlorophanus. *Z. Naturforsch. B* **11b,** 513–524 (1956).

29. Sanes, J. R. & Zipursky, L. Design principles of insect and vertebrate visual systems. *Neuron* **66,** 15–36 (2010).

30. Fischbach, K. F. & Dittrich, A. P. M. The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure. *Cell Tissue Res.* **258,** 441–475 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

METHODS

Flies. Flies were raised on standard cornmeal-agar medium at a 12 h light/12 h dark cycle, 25 °C and 60% humidity. We used female experimental flies, 1 day after eclosion. Two effector strains carrying the *white* gene with multiple insertions of *UAS-shi^{ts}* on the third chromosome or a single insertion of *UAS-Kir2.1-GFP* on the second chromosome were used for blocking experiments. Heterozygous control and experimental flies were obtained by crossing the respective *GAL4* driver and *UAS*-effector strains. For the *shibire* experiments, control and experimental flies had identical genotype and were raised at 18 °C (permissive temperature). Experimental flies were shifted for 1 h to 37 °C (restrictive temperature) directly before the experiment and recorded at room temperature. No recovery of the block was detected within the time of recording (see Supplementary Information). For the *Kir2.1* experiments, experimental and control flies were raised at 30 °C to boost the expression level of *Kir2.1*. Three sets of *white*[+] control flies were used for the experiments shown in Fig. 3e: *UAS-Kir2.1-GFP* ($n = 3$), L1a ($n = 5$) and L2c ($n = 4$). To restore photoreceptor input to L1 or L2, *DB331-GAL4, UAS-mCD8-TnXL-8aa* (for highlighting tangential cells[2]); *UAS-ort; ort^{US2515}* was crossed to *c202a-GAL4* (L1b); *ninaE^1, ort^1* or *21D-GAL4* (L2b), *ort^1* (ref. 11). *DB331-GAL4, UAS-mCD8-TnXL-8aa/+; UAS-ort/+; ort^{US2515}/+* served as a positive control, whereas *DB331-GAL4, UAS-mCD8-TnXL-8aa/+; UAS-ort/+; ort^{US2515}/ninaE^1, ort^1* and *DB331-GAL4, UAS-mCD8-TnXL-8aa/+; c202a-GAL4* (L1b)/+; *ort^{US2515}/ninaE^1, ort^1* served as negative controls. The driver lines had the following genotypes: L1a, *vGlut-dVP16AD/CyO; ortC2-GAL4DBD/TM3* (ref. 18), provided by C.-H. Lee; L1b, *c202-GAL4* (ref. 11), provided by J. Rister; L2a, *ortC1-3-GAL4AD; +; cha-GAL4DBD* (ref. 18), provided by C.-H. Lee; L2b, *21D-GAL4* (ref. 11), provided by J. Rister; L2c, *ortC3-GAL4* (ref. 18), provided by C.-H. Lee; L1L2, *6298-GAL4* (ref. 11), provided by J. Rister.

Preparation. Flies were anaesthetized on ice and waxed on a Plexiglas holder using bee wax. The dissection of the fly cuticle and exposure of the lobula plate was performed as in ref. 2. A Ringer's-solution-filled cleaning electrode (tip ~4 μm) was used to remove the extracellular matrix and to expose the somata of lobula plate tangential cells for recording. These somata were recognized either by expression of a fluorescent marker (see above) or by their location next to a prominent tracheal branch.

Whole-cell recording. VS and HS cell somata covered by Ringer's solution were approached with a patch electrode filled with a red fluorescent dye (intracellular solution as in ref. 2). Recordings in rescue experiments were performed with fluorescently targeted neurons as in ref. 3; recordings in blocking experiments were established under high-contrast optics using a ×40 water immersion objective (LumplanF, Olympus), a Zeiss Microscope (Axiotech vario 100, Zeiss) and illumination (100 W fluorescence lamp, heat mirror, neutral density filter OD 0.3; all from Zeiss). To enhance tissue contrast, we used two polarization filters, one located as an excitation filter and the other as an emission filter, with slight deviation on their polarization plane. For eye protection, we additionally used a 420-nm LP filter on the light path. For further details of the set-up, see ref. 2.

Neurobiotin dye fill and staining. Flies expressing mCD8-GFP in L1 or L2 (L1a or L2b) were decapitated. The heads were fixed and lamina cells were injected with a 5 mM Alexa Fluor 568 (A10442; Invitrogen) and 2% neurobiotin (Vector Labs) solution as described in ref. 3. Access to the lamina monopolar cell somata was possible after removal of the fly's retina. A single lamina cell was injected per brain. Neurobiotin was detected by coupling to Streptavidin–Alexa Fluor 568 conjugate (1:100, Invitrogen)[2].

Immunohistochemistry. Female flies 3–5 days after eclosure were dissected. Their heads were fixed in freshly prepared 4% paraformaldehyde in PBT (overnight at 4 °C). Subsequently, the heads were washed for 45–60 min in PBT and mounted in 7% agarose. Agarose blocks containing a single fly head were sectioned at 50 μm using a vibratome (Leica VT 1000S). The sections were immediately treated with 2% sodium borohydrate (806372, MERCK) for 20 min at room temperature to reduce the auto-fluorescence. After proper washing in PBT for 45–60 min, the sections were further incubated in PBT including 2% normal goat serum (50-062Z, Invitrogen) and subsequently in primary antibodies (1:200, overnight at 4 °C). Antibodies were removed by several washing steps (5 × 20 min in PBT) and secondary antibodies were added (1:200, overnight at 4 °C). A 5 × 20 min washing protocol (PBT) was followed by final washing steps in PBS (5 × 20 min). The stained tissues were mounted in Vectashield (Vector Laboratories) and analysed by confocal microscopy. The following primary and secondary antibodies were used in the present study: rat anti-mCD8 (MCD0800, Caltag laboratories), mouse anti-Dlg (Developmental Studies Hybridoma Bank), rabbit anti-ShakB antibodies (J. Bacon), Alexa Fluor 568 goat anti-rat-IgG (A11077, Molecular Probes), Alexa Fluor 568 goat anti-rabbit-IgG (A11011, Molecular Probes) and Alexa Fluor 488 goat anti-mouse-IgG (A11001, Molecular Probes). Female experimental flies of the following genotypes were used for immunohistochemistry: +/+; *vGlut-dVP16AD/UAS-mCD8-DsRed; ortC2-GAL4DBD/TM3* (for L1a), +/+; *c202-GAL4/UAS-mCD8-DsRed; +/+* (for L1b), *ortC1-3-GAL4AD/+; UAS-mCD8-DsRed/+; cha-GAL4DBD/+* (for L2a), +/+; *UAS-mCD8-DsRed/+; 21D-GAL4/+* (for L2b), +/+; *UAS-mCD8-DsRed/+; ortC3-GAL4/+* (for L2c), and +/+; *6298-GAL4/UAS-mCD8-DsRed; +/+* (for L1L2).

Confocal microscopy. Serial optical sections were taken at 0.5-μm intervals with 1,024 × 1,024 pixel resolution using a confocal microscope (Leica TCS-NT) and an oil-immersion ×40- (NA = 1.25) Plan-Apochromat objective or a ×40 water-immersion objective (LUMPlanF, Olympus). The individual confocal stacks were analysed using Image J (NIH). The size, contrast and brightness of the resulting images were adjusted with Photoshop CS (Adobe Systems).

Visual stimulation. A custom-built LED arena covered ~170° (1.4° resolution) of the horizontal and ~100° (1.4° resolution) of the vertical visual field of the fly, allowing refresh rates of up to 600 Hz with 16 intensity levels. The spectral peak of the LEDs was at 568 nm and the luminance range of the stimuli was between 0.5–8 cd m^{-2} (for further details see ref. 2). Two types of visual stimuli were used: the moving grating consisted of either a square-wave (Fig. 2) or a sine-wave (Fig. 3) pattern with a spatial wavelength of 22° moving at 22° s^{-1}. The moving ON or OFF edge (Fig. 4) consisted of an edge of either polarity moving at 44° s^{-1}.

Data analysis. Data were acquired and analysed using the data acquisition and analysis toolboxes of Matlab (The Mathworks). The contrast was calculated as $(I_{max} - I_{min})/(I_{max} + I_{min})$ with an absolute I_{min} and I_{max} of 0.5 and 8 cd m^{-2}, respectively. The responses were defined as the difference between the average membrane voltage during the 2-s stimulation period and the 500-ms average potential before stimulation.

LETTER

Reducing excessive GABA–mediated tonic inhibition promotes functional recovery after stroke

Andrew N. Clarkson[1]*†, Ben S. Huang[1,2]*, Sarah E. MacIsaac[1], Istvan Mody[1,2,3] & S. Thomas Carmichael[1]

Stroke is a leading cause of disability, but no pharmacological therapy is currently available for promoting recovery. The brain region adjacent to stroke damage—the peri-infarct zone—is critical for rehabilitation, as it shows heightened neuroplasticity, allowing sensorimotor functions to re-map from damaged areas[1–3]. Thus, understanding the neuronal properties constraining this plasticity is important for the development of new treatments. Here we show that after a stroke in mice, tonic neuronal inhibition is increased in the peri-infarct zone. This increased tonic inhibition is mediated by extrasynaptic GABA$_A$ receptors and is caused by an impairment in GABA (γ-aminobutyric acid) transporter (GAT-3/GAT-4) function. To counteract the heightened inhibition, we administered _in vivo_ a benzodiazepine inverse agonist specific for α5-subunit-containing extrasynaptic GABA$_A$ receptors at a delay after stroke. This treatment produced an early and sustained recovery of motor function. Genetically lowering the number of α5- or δ-subunit-containing GABA$_A$ receptors responsible for tonic inhibition also proved beneficial for recovery after stroke, consistent with the therapeutic potential of diminishing extrasynaptic GABA$_A$ receptor function. Together, our results identify new pharmacological targets and provide the rationale for a novel strategy to promote recovery after stroke and possibly other brain injuries.

Stroke is a major source of disability, confining one-third of stroke survivors to nursing homes or institutional settings[4]. Recent studies have shown that the brain has a limited capacity for repair after stroke. Neural repair after stroke involves re-mapping of cognitive functions in tissue adjacent to or connected with the stroke[5,6]. Functional recovery in this peri-infarct tissue involves changes in neuronal excitability that alter the brain's representation of motor and sensory functions. Stimulation of peri-infarct cortex enhances local neuronal excitability through a process that involves long-term potentiation, alters sensorimotor maps and improves the use of affected limbs[5–8]. The inhibitory neurotransmitter GABA is critical for cortical plasticity and sensory mapping. Altering GABAergic transmission changes sensory maps during the critical period of cortical development[9] and produces rapid alterations in adult cortical maps that resemble changes occurring after stroke[10,11]. Alterations in cortical maps through blockade of GABAergic signalling are associated with fundamental changes in cellular excitability including long-term potentiation[12]. In a similar manner to normal cortical plasticity, GABAergic mechanisms may mediate changes in neuronal excitability that have a central role in functional recovery of peri-infarct cortex after stroke.

Cortical GABAergic signalling through GABA$_A$ receptors is divided into synaptic (phasic) and extrasynaptic (tonic) components. Tonically active extrasynaptic GABA$_A$ receptors set an excitability threshold for neurons[13,14]. Extrasynaptic GABA$_A$ receptors primarily consist of α5- or δ-subunit-containing receptors[13,14]. Pharmacological and genetic knockdown of α5-GABA$_A$ receptors enhance long-term potentiation and improve performance on learning and memory tasks[15,16].

The selective effects of extrasynaptic GABA$_A$ receptors on cellular excitability and plasticity, and the evidence that changes in neuronal excitability underlie functional reorganization in peri-infarct cortex, indicate that this system may have a role in recovery after stroke. We find that stroke increases tonic GABAergic transmission in peri-infarct cortex and that dampening this tonic inhibition produces an early and robust gain of motor recovery after stroke (Supplementary Fig. 1, schematic summary).

We examined neuronal excitability in the peri-infarct cortex of mice during the period of recovery and reorganization after a photothrombotic stroke to forelimb motor cortex. Whole-cell voltage-clamp recordings in _in vitro_ brain slices prepared at 3, 7 and 14 days after stroke (Fig. 1a) showed a significant increase in GABA$_A$-receptor-mediated tonic inhibition (I_{tonic}) in layer 2/3 pyramidal neurons, compared to neurons from sham controls (control: 8.05 ± 0.80 pA pF^{-1}, $n = 24$, versus after stroke: 13.6 ± 1.41 pA pF^{-1}, $n = 45$, Mann–Whitney U-test, $P < 0.05$; Fig. 1b–d). I_{tonic} remained raised from 3 to 14 days after stroke (Supplementary Fig. 2a). The mean phasic excitation remained unchanged over the 2-week period after stroke
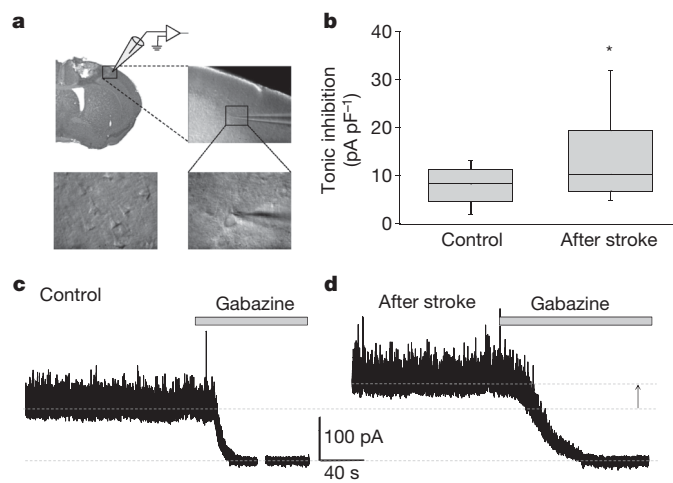
Figure 1 | Increased tonic inhibition in peri-infarct cortex. a, Images showing the peri-infarct recording site. Whole-cell patch-clamp recordings were made from post-stroke brain slices, within 200 µm of the infarct (top left; image under 2× magnification), from layer-2/3 (top right; 10×) pyramidal neurons (bottom panels; 40×). **b,** Box plot (boxes, 25–75%; whiskers, 10–90%; lines, median) showing significantly increased tonic inhibition in peri-infarct cortex (asterisk, $P < 0.05$; see Supplementary Fig. 2 for further analyses). **c, d,** Representative traces showing tonic inhibitory currents in control and peri-infarct neurons, respectively. Tonic currents were revealed by the shift in holding currents after blocking all GABA$_A$ receptors with gabazine (>100 µM). Cells were voltage clamped at +10 mV.

[1]Department of Neurology, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. [2]Interdepartmental PhD Program for Neuroscience, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. [3]Department of Physiology, The David Geffen School of Medicine at UCLA, 635 Charles Young Drive South, Los Angeles, California 90095, USA. †Present address: Departments of Psychology and Anatomy and Structural Biology, University of Otago, PO Box 913, Dunedin 9013, New Zealand. *These authors contributed equally to this work.

(Supplementary Fig. 3a, c). The mean phasic inhibition was unchanged except for a transient decrease at 7 days after stroke (Supplementary Fig. 3b, d). The resting membrane and GABA reversal potentials were both unchanged (Supplementary Fig. 3e, f).

Tonic inhibition is effectively controlled by the degree of extracellular GABA uptake through neuronal and astrocytic GATs[14]. We applied a GAT-1-selective antagonist, NO-711 (10 μM), and found a significantly greater effect (percentage increase in I_{tonic} after GAT blockade) in post-stroke neurons ($94.0 \pm 16.3\%$, $n = 10$) than in controls ($34.3 \pm 11.4\%$, $n = 6$; $P < 0.05$; Fig. 2a). Co-application of NO-711 and the GAT-3/GAT-4-selective antagonist SNAP-5114 (40 μM) produced a substantial increase in I_{tonic} in controls ($300.6 \pm 46.0\%$, $n = 4$; Fig. 2a), revealing the synergistic actions of GATs in the cortex as previously proposed[17]. In post-stroke neurons, co-application only produced an effect ($110.7 \pm 32.0\%$, $n = 5$) similar to GAT-1 blockade alone ($P = 0.68$; Fig. 2a), indicating a dysfunction in GAT-3/GAT-4 after stroke. Sequential blockade of the two GATs confirmed the impairment after stroke, as peri-infarct I_{tonic} showed no further response to GAT-3/GAT-4 blockade after the initial GAT-1 block, in contrast to responses shown in controls (Fig. 2b, c). This effect was not due to receptor saturation, as I_{tonic} showed a further increase in response to a raised concentration of GABA under GAT blockade (Supplementary Fig. 4a). Western-blot analysis confirmed a reduced GAT-4 level in peri-infarct cortex, whereas GAT-1 levels were unchanged (Supplementary Fig. 5).

We proposed that the chronically raised tonic inhibition in the peri-infarct region may antagonize the neuronal plasticity required for functional recovery after stroke. Therefore, we tested whether reducing the excessive tonic inhibition would improve functional recovery. Of the two GABA_A receptor subtypes shown to underlie tonic inhibition in cortical neurons, α5-GABA_A receptors can be antagonized specifically by L655,708, a benzodiazepine inverse agonist[16], whereas no specific antagonist exists for δ-GABA_A receptors. L655,708 (100 nM) decreased I_{tonic} in control neurons by $-13.3 \pm 5.2\%$ ($n = 4$), but produced a significantly greater decrease in post-stroke neurons ($-30.0 \pm 4.1\%$, $n = 13$; $P < 0.05$; Fig. 2d, e), which reverted I_{tonic} back to control level (control: see earlier versus after stroke plus L655,708: 140.8 ± 18.5 pA, $n = 13$; $P = 0.702$; Fig. 2f). L655,708 produced only minimal effects on phasic inhibitory currents in both post-stroke and control conditions (Supplementary Fig. 4b).

Next we tested the effects of reducing tonic inhibition on functional recovery after stroke, using measures of forelimb and hindlimb motor control. Stroke produced an increase in the number of foot faults in the grid-walking task, and a decrease in forelimb symmetry in the cylinder task from 7 days after stroke. Chronic treatment with L655,708 starting 3 days after stroke resulted in a dose-dependent maximal gain of function beginning from 7 days after stroke in both tasks ($P < 0.001$; Fig. 3a–c). Acute treatment with L655,708 just before behavioural testing had a minimal effect on stroke recovery (Supplementary Fig. 7). To assess the necessity of long-term administration, we discontinued
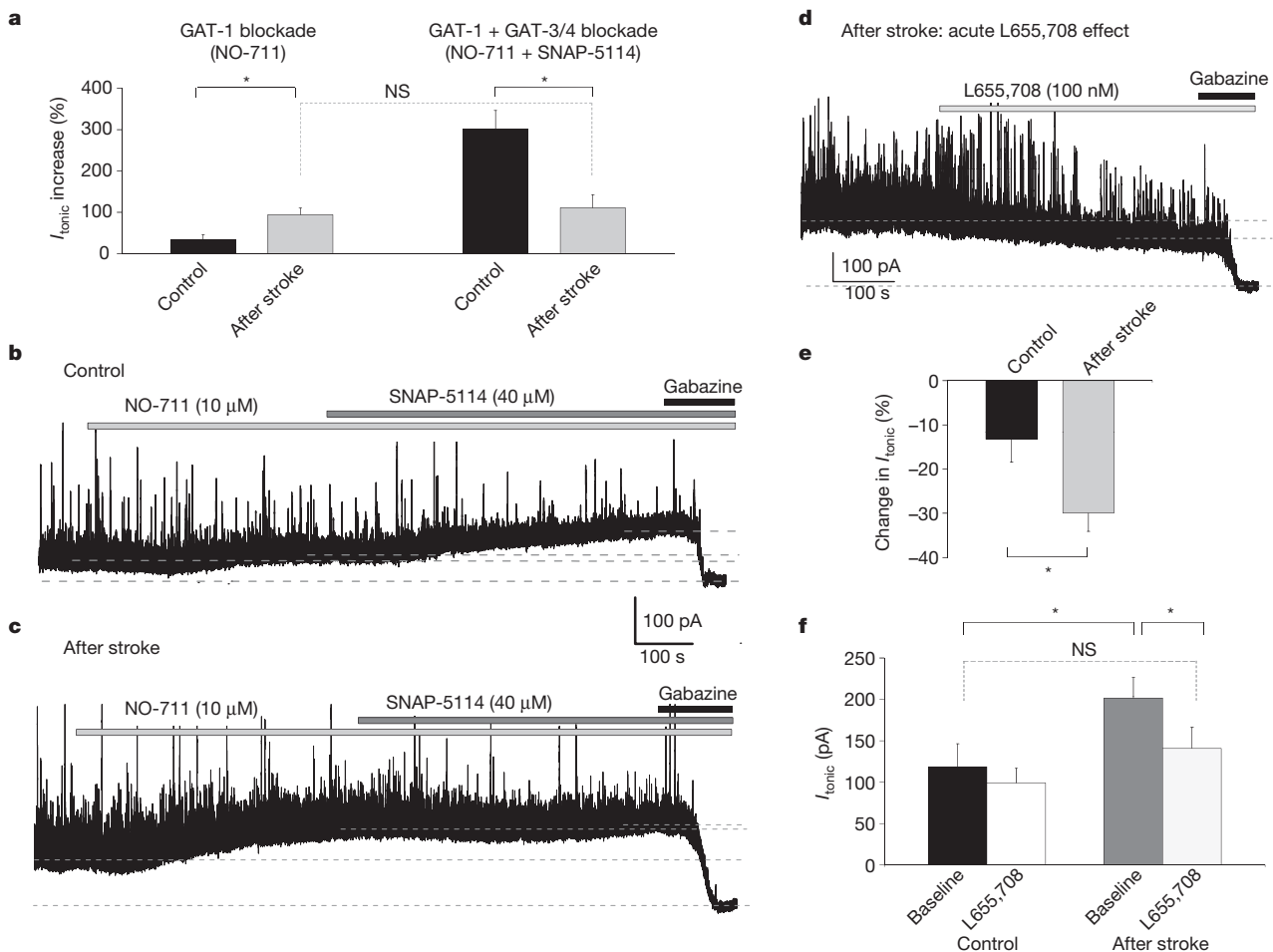
Figure 2 | Impairment in GABA transport and effect of blocking α5-GABA_A receptors after stroke. a, Blocking GAT-1 (NO-711) produced a higher percentage increase in I_{tonic} after stroke; combined blockade of GAT-1 and GAT-3/GAT-4 (NO-711 plus SNAP-5114) produced a substantial I_{tonic} increase in controls but only an increase equivalent to blocking GAT-1 alone after stroke. **b, c**, I_{tonic} in sequential drug applications. Note the lack of response to SNAP-5114 application in the post-stroke slice. **d**, L655,708 reduced I_{tonic}. **e**, L655,708 significantly decreased I_{tonic} after stroke. **f**, Drug treatment reverted post-stroke I_{tonic} to near-control level. *, $P < 0.05$; NS, not significant; bar graphs represent mean ± s.e.m.

L655,708 treatment after 2 weeks and found a decrease in functional gains, although these mice still performed better than vehicle-treated stroke controls (Supplementary Fig. 6).

To corroborate further the role of reduced tonic inhibition in enhancing stroke recovery, we tested mice with deletions of either $\alpha5$- or δ-subunit-containing $GABA_A$ receptors ($Gabra5^{-/-}$ and $Gabrd^{-/-}$; Methods)[18]. $Gabra5^{-/-}$ animals showed significantly better motor recovery after stroke, comparable to L655,708-treated wild-type animals (Fig. 3d–f). In addition, $Gabra5^{-/-}$ animals showed a significant reduction in hindlimb foot faults (Fig. 3e). $Gabrd^{-/-}$ animals also showed significant improvements in motor recovery (Fig. 3d–f), but to a lesser extent than the $Gabra5^{-/-}$ mice. Thus, modulation of $\alpha5$-$GABA_A$ receptors affords greater functional gains in motor recovery than δ-$GABA_A$ receptors, and genetic removal of $\alpha5$-$GABA_A$ receptors produces a more widespread increase in motor recovery than pharmacological antagonism. Administration of L655,708 to $Gabrd^{-/-}$ mice produced an even greater recovery, confirming the beneficial effect of reducing peri-infarct tonic inhibition (Fig. 3a, c).

Low/sub-seizure dosing of picrotoxin (PTX; 0.1 mg kg^{-1}, administered intraperitoneally (i.p.)), a use-dependent $GABA_A$ receptor antagonist, enhances learning and memory in transgenic mouse models of Alzheimer's and other cognitive impairments by reversing

an increased GABAergic inhibitory tone, acting at both synaptic and extrasynaptic $GABA_A$ receptors[19,20]. The pharmacological effects of PTX on reducing phasic and tonic inhibition were not altered after stroke (Supplementary Table 2). PTX given to animals from 3 days after stroke resulted in a significant gain of forelimb function on the grid-walking task compared to vehicle-treated stroke controls ($P < 0.05$; Supplementary Fig. 9a). No significant changes were observed in hindlimb function or forelimb asymmetry (Supplementary Fig. 9b, c). Combined L655,708 and PTX treatment showed similar initial functional gains compared to stroke and L655,708 alone; however, prolonged PTX and L655,708 treatment produced a deterioration in motor function such that the performance progressively worsened at late periods after stroke (Supplementary Fig. 9). These data indicate that increasing cortical excitability too far or reducing phasic inhibition negatively affects functional recovery.

An important element in stroke treatment is the timing of drug delivery. $GABA_A$ receptor agonists administered at the time of stroke decrease stroke size[20]. Therefore, dampening tonic inhibition too early after stroke may produce an opposite effect, that is, increased cell death. To test this, we assessed stroke volume at 7 days after stroke, in animals treated with 1) vehicle; 2) L655,708 from stroke onset; and 3) L655,708 from day 3 after stroke. Stroke volumes were similar between mice
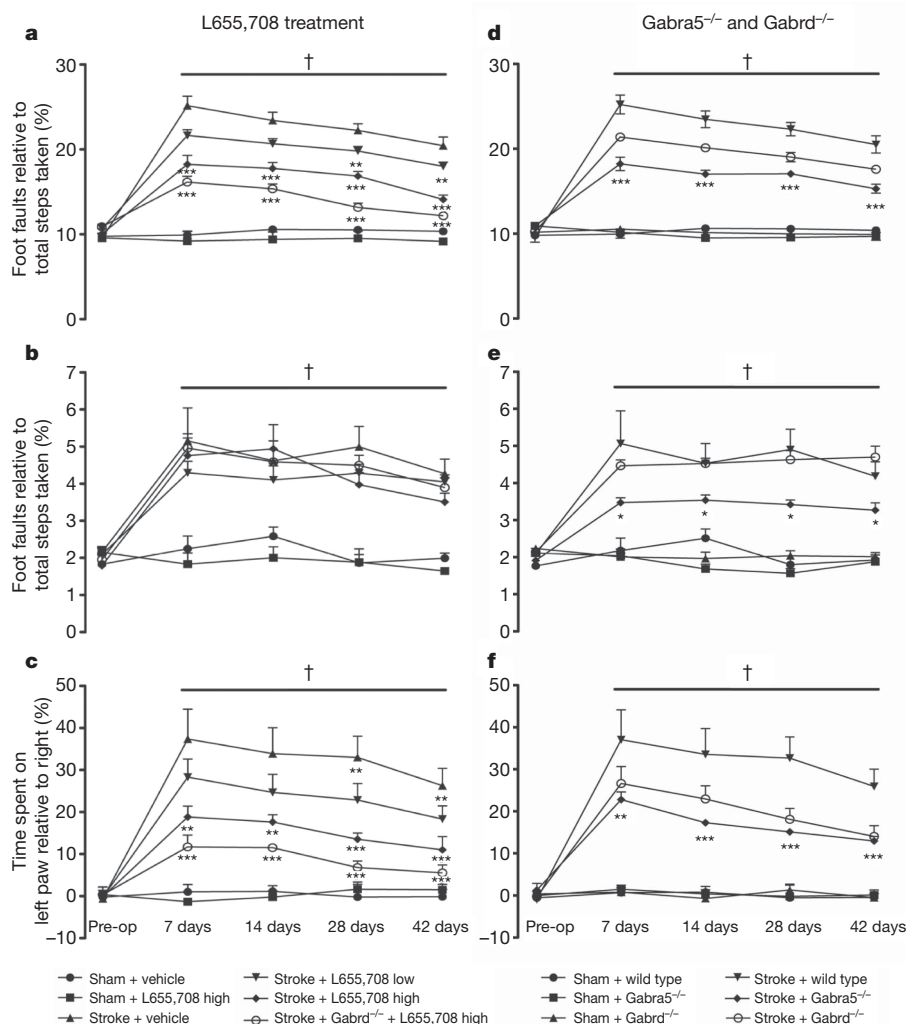
Figure 3 | Behavioural recovery after stroke with L655,708 treatment and in Gabra5$^{-/-}$ and Gabrd$^{-/-}$ animals. a–c, L655,708 treatment starting from 3 days after stroke resulted in a dose-dependent improvement in functional recovery after stroke. **d–f,** Gabra5$^{-/-}$ and Gabrd$^{-/-}$ mice also showed decreased motor deficits after stroke. Functional recovery was assessed with forelimb (**a, d**) and hindlimb foot faults (**b, e**), and on forelimb asymmetry (**c, f**). Low-dose L655,708 = 200 µg kg^{-1} day^{-1} per animal; high-dose L655,708 = 400 µg kg^{-1} day^{-1} per animal. Pre-op, pre-operation. Data are mean ± s.e.m. †, $P \leq 0.001$ stroke plus vehicle versus sham; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$ versus stroke plus vehicle.
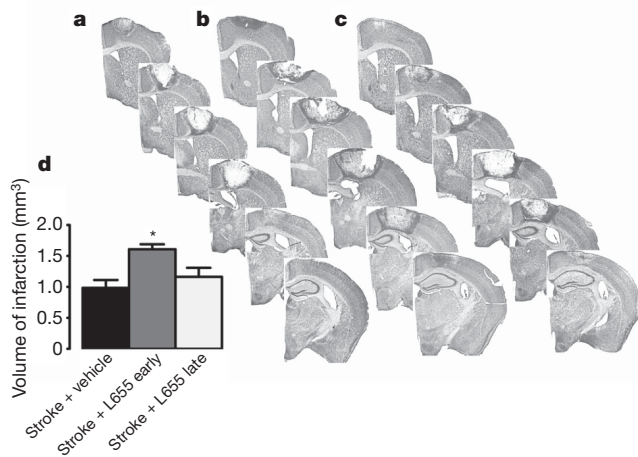
Figure 4 | Inflection point in L655,708 treatment effect on infarct size.
a–d, Representative Nissl-stained sections at 7 days after stroke from stroke plus vehicle treatment (**a**), stroke plus L655,708 treatment starting at the time of stroke (**b**) and stroke plus L655,708-treatment starting from 3 days post-insult (**c**). Quantification of the stroke volume is shown in panel **d**. Data are mean ± s.e.m. for $n = 4$ per group; $* = P \leq 0.05$.

treated with vehicle and L655,708 from day 3 (Fig. 4). In contrast, stroke volume was significantly increased in animals treated with L655,708 from stroke onset ($P < 0.05$; Fig. 4). These data indicate a critical timeframe for therapeutically dampening tonic inhibition after stroke: reduction too early would exacerbate stroke damage, whereas delaying treatment by 3 days would promote functional recovery without altering stroke size. Genetic deletion of α5- or δ-GABA$_A$ receptors did not affect infarct size or neuronal number in peri-infarct cortex (Supplementary Fig. 8). Unlike pharmacological antagonism of α5-GABA$_A$-receptor-mediated inhibition, in Gabra5$^{-/-}$ and Gabrd$^{-/-}$ mice, the genomic absence of one of the extrasynaptic GABA$_A$ receptors may trigger compensatory upregulation of the other receptor[13], thus obscuring their roles in neuroprotection immediately after stroke.

Current therapies that promote functional recovery after stroke are limited to physical rehabilitation[4]. Here, by identifying an excessive tonic inhibition after stroke, we have found promising new targets for pharmacological interventions to promote recovery. The increase in tonic inhibition in cortical pyramidal neurons occurs during precisely the same time period important for cortical map plasticity and recovery[1–3]. Alterations in other aspects of cortical signalling have also been described during this period, including altered GABA$_A$ receptor subunits, glutamate receptor expression and neuronal network properties[21–24]. Protein levels of GAT-1 were shown to be decreased in peri-infarct cortex in some rodent stroke models, and reactive astrocytes show reduced uptake of other neurotransmitters[24]. However, there are conflicting data on GABA$_A$ receptor levels after stroke[23–26]. We found a decreased protein level and compromised function of GAT-3/GAT-4 in peri-infarct cortex. The increase in tonic inhibition may curtail cortical plasticity and spontaneous recovery after stroke, and is consistent with tonic GABAergic inhibition exerting a causal role in limiting motor recovery in stroke.

Non-selectively decreasing GABAergic tone facilitates neuronal plasticity in genetic models of cognitive diseases[19,20]. We show for the first time, to our knowledge, that antagonizing a raised tonic inhibition enhances motor recovery after stroke, consistent with the idea that molecular and cellular events of neuronal plasticity are dampened in the peri-infarct zone, and promoting this plasticity facilitates functional recovery. Together, our results have identified novel pharmacological targets and provide a rational basis for developing future therapies to promote recovery after stroke and possibly other brain injuries.

METHODS SUMMARY

Photothrombotic model of focal stroke. Focal stroke was induced by photothrombosis in adult male C57BL/6 mice (aged 2–4 months) as described previously[27].
Slice preparation for electrophysiology. Following decapitation, brains were rapidly removed and placed into an N-methyl-D-glucamine (NMDG)-based cutting solution to enhance neuronal viability[28]. Coronal slices (350 μm) were cut and transferred to an interface-style chamber containing artificial cerebrospinal fluid (ACSF) as previously described[13]. Recordings were made from intact peri-infarct cortical layer-2/3 pyramidal neurons and analysed as previously described[13].
In vivo drug administration. L655,708 was dissolved in dimethylsulphoxide (DMSO) and then diluted 1:1 in 0.9% saline. L655,708-filled ALZET-1002 pumps were implanted at 3 days after stroke and replaced every two weeks. In acute administration studies, 5 mg kg^{-1} L655,708 was administered i.p. 30 min before testing. The concentration in one minipump, 5 mM, delivers a 200 μg kg^{-1} day^{-1} dose in mice. With one or two minipumps implanted, this provides a dose escalation. PTX (0.1 mg kg^{-1} i.p. twice a day) starting 3 days after stroke was administered alone or together with L655,708.
Behavioural analysis. Mice were videotaped during walking and exploratory behaviour in the grid-walking and cylinder/rearing tasks, and were tested at approximately the same time each day during the nocturnal period[29]. Baseline behavioural measurements were obtained one week before surgery. Post-stroke animals were assessed at weeks 1, 2, 4 and 6.
Infarct-size measurement. For the histological assessment of infarct size, brains were processed at 7 days after stroke using cresyl violet as previously described[30].
Statistical analyses. All data are expressed as mean ± s.e.m. For electrophysiological comparisons between control and post-stroke animals, the Mann–Whitney non-parametric test was used. For multiple comparisons across days after stroke, one-way analysis of variance (ANOVA) and Newman–Keuls' multiple pair-wise comparisons for post-hoc comparisons were used. For behavioural testing, differences between treatment groups were analysed using two-way ANOVA with repeated measures and Newman–Keuls' multiple pair-wise comparisons. The level of significance was set at $P < 0.05$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cramer, S. C. Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. *Ann. Neurol.* **63,** 272–287 (2008).
2. Brown, C. E., Aminoltejari, K., Erb, H., Winship, I. R. & Murphy, T. H. In vivo voltage-sensitive dye imaging in adult mice reveals that somatosensory maps lost to stroke are replaced over weeks by new structural and functional circuits with prolonged modes of activation within both the peri-infarct zone and distant sites. *J. Neurosci.* **29,** 1719–1734 (2009).
3. Dijkhuizen, R. M. *et al.* Correlation between brain reorganization, ischemic damage, and neurologic status after transient focal cerebral ischemia in rats: a functional magnetic resonance imaging study. *J. Neurosci.* **23,** 510–517 (2003).
4. Dobkin, B. H. Training and exercise to drive poststroke recovery. *Nature Clin. Pract. Neurol.* **4,** 76–85 (2008).
5. Carmichael, S. T. Cellular and molecular mechanisms of neural repair after stroke: making waves. *Ann. Neurol.* **59,** 735–742 (2006).
6. Nudo, R. J. Mechanisms for recovery of motor function following cortical damage. *Curr. Opin. Neurobiol.* **16,** 638–644 (2006).
7. Alonso-Alonso, M., Fregni, F. & Pascual-Leone, A. Brain stimulation in poststroke rehabilitation. *Cerebrovasc. Dis.* **24** (suppl. 1), 157–166 (2007).
8. Di Lazzaro, V. *et al.* Motor cortex plasticity predicts recovery in acute stroke. *Cereb. Cortex* **20,** 1523–1528 (2010).
9. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6,** 877–888 (2005).
10. Donoghue, J. P., Suner, S. & Sanes, J. N. Dynamic organization of primary motor cortex output to target muscles in adult rats. II. Rapid reorganization following motor nerve lesions. *Exp. Brain Res.* **79,** 492–503 (1990).
11. Foeller, E., Celikel, T. & Feldman, D. E. Inhibitory sharpening of receptive fields contributes to whisker map plasticity in rat somatosensory cortex. *J. Neurophysiol.* **94,** 4387–4400 (2005).
12. Hess, G., Aizenman, C. D. & Donoghue, J. P. Conditions for the induction of long-term potentiation in layer II/III horizontal connections of the rat motor cortex. *J. Neurophysiol.* **75,** 1765–1778 (1996).
13. Glykys, J. & Mody, I. Hippocampal network hyperactivity after selective reduction of tonic inhibition in GABA$_A$ receptor α5 subunit-deficient mice. *J. Neurophysiol.* **95,** 2796–2807 (2006).
14. Walker, M. C. & Semyanov, A. Regulation of excitability by extrasynaptic GABA(A) receptors. *Results Probl. Cell Differ.* **44,** 29–48 (2008).
15. Collinson, N. *et al.* Enhanced learning and memory and altered GABAergic synaptic transmission in mice lacking the α5 subunit of the GABA$_A$ receptor. *J. Neurosci.* **22,** 5572–5580 (2002).

16. Atack, J. R. *et al.* L-655,708 enhances cognition in rats but is not proconvulsant at a dose selective for α5-containing GABA_A receptors. *Neuropharmacology* **51**, 1023–1029 (2006).
17. Keros, S. & Hablitz, J. J. Subtype-specific GABA transporter antagonists synergistically modulate phasic and tonic GABA_A conductances in rat neocortex. *J. Neurophysiol.* **94**, 2073–2085 (2005).
18. Glykys, J. & Mody, I. Activation of GABA_A receptors: views from outside the synaptic cleft. *Neuron* **56**, 763–770 (2007).
19. Yoshiike, Y. *et al.* GABA_A receptor-mediated acceleration of aging-associated memory decline in APP/PS1 mice and its pharmacological treatment by picrotoxin. *PLoS ONE* **3**, e3029 (2008).
20. Cui, Y. *et al.* Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* **135**, 549–560 (2008).
21. Ginsberg, M. D. Neuroprotection for ischemic stroke: past, present and future. *Neuropharmacology* **55**, 363–389 (2008).
22. Que, M. *et al.* Changes in GABA(A) and GABA(B) receptor binding following cortical photothrombosis: a quantitative receptor autoradiographic study. *Neurosci. Lett.* **93**, 1233–1240 (1999).
23. Redecker, C., Luhmann, H. J., Hagemann, G., Fritschy, J. M. & Witte, O. W. Differential downregulation of GABA_A receptor subunits in widespread brain regions in the freeze-lesion model of focal cortical malformations. *J. Neurosci.* **20**, 5045–5053 (2000).
24. Frahm, C. *et al.* Regulation of GABA transporter mRNA and protein after photothrombotic infarct in rat brain. *J. Comp. Neurol.* **478**, 176–188 (2004).
25. Neumann-Haefelin, T. *et al.* Immunohistochemical evidence for dysregulation of the GABAergic system ipsilateral to photochemically induced cortical infarcts in rats. *Neuroscience* **87**, 871–879 (1998).
26. Kharlamov, E. A., Downey, K. L., Jukkola, P. I., Grayson, D. R. & Kelly, K. M. Expression of GABA_A receptor α1 subunit mRNA and protein in rat neocortex following photothrombotic infarction. *Brain Res.* **1210**, 29–38 (2008).
27. Lee, J. K., Kim, J. E., Sivula, M. & Strittmatter, S. M. Nogo receptor antagonism promotes stroke recovery by enhancing axonal plasticity. *J. Neurosci.* **24**, 6209–6217 (2004).
28. Tanaka, Y., Furuta, T., Yanagawa, Y. & Kaneko, T. The effects of cutting solutions on the viability of GABAergic interneurons in cerebral cortical slices of adult mice. *J. Neurosci. Methods* **171**, 118–125 (2008).
29. Baskin, Y. K., Dietrich, W. D. & Green, E. J. Two effective behavioral tasks for evaluating sensorimotor dysfunction following traumatic brain injury in mice. *J. Neurosci. Methods* **129**, 87–93 (2003).
30. Ohab, J. J., Fleming, S., Blesch, A. & Carmichael, S. T. A neurovascular niche for neurogenesis after stroke. *J. Neurosci.* **26**, 13007–13016 (2006).

METHODS

Photothrombotic model of focal stroke. Under isoflurane anaesthesia (2–2.5% in a 70% N_2O/30% O_2 mixture), 2–4 month-old adult C57BL/6 (Charles River) male mice were placed in a stereotactic apparatus, the skull exposed through a midline incision, cleared of connective tissue and dried. A cold light source (KL1500 LCD, Zeiss) attached to a $40\times$ objective giving a 2-mm diameter illumination was positioned 1.5 mm lateral from Bregma, and 0.2 ml of Rose Bengal solution (Sigma; $10\,\mathrm{g\,l^{-1}}$ in normal saline, i.p.) was administered[31]. After 5 min, the brain was illuminated through the intact skull for 15 min. Rose Bengal produces singlet oxygen under light excitation, which damages and occludes vascular endothelium, resulting in focal cortical stroke under the region of illumination (Fig. 4), circumscribed by peri-infarct tissue with normal neuronal cell number (Supplementary Fig. 8). Two-to-four-month-old adult male Gabra5$^{-/-}$ and Gabrd$^{-/-}$ mice[32] received a stroke as described earlier. These mice had been backcrossed to C57BL/6 mice in excess of 15 generations, and were compared in behavioural studies to wild-type C57BL/6. Body temperature was maintained at $36.9 \pm 0.4\,°C$ with a heating pad throughout the operation and did not vary by drug or genetic condition. This stroke method produces a small stroke in the mouse forelimb region of the motor cortex (Fig. 4). Sample size was ten per group for Gabra5$^{-/-}$ and Gabrd$^{-/-}$ in stroke/behavioural studies. Sample size was eight per group for each condition in dosing of L655,708 (Fig. 3).

Blood pressure (systolic and diastolic) and heart rate were measured in separate cohorts of wild-type (C57BL/6) mice, with or without L655,708 administration via ALZET minipumps from 3 days after stroke, and in Gabra5$^{-/-}$ and Gabrd$^{-/-}$ mice, before, during and after stroke, using a standard non-invasive tail-cuff method (Coda). There were no significant differences in heart rate or blood pressure by treatment or genotype (Supplementary Table 1). All studies in this manuscript complied with the Stroke Therapy Academic Industry Roundtable (STAIR) criteria[33,34] for stroke investigations in measuring physiological parameters, monitoring treatment effects for at least one month, analysing treatment effects blinded to conditions, using dose–response studies and use of a drug administration route with blood–brain-barrier penetration.

Whole-cell voltage-clamp electrophysiology. Slices were submerged in the recording chamber and continuously perfused ($5–8\,\mathrm{ml\,min^{-1}}$) with oxygenated ACSF ($32–34\,°C$). Visualized patch-clamp recordings from layer-2/3 pyramidal neurons were performed at $40\times$ using infrared oblique-illumination (Leica DM-LFS; Hamamatsu CCD camera C3077-78).

Control recordings were made from cells of sham-operated animals at similar locations as those recorded in post-stroke animals. Microelectrodes ($3–5\,\mathrm{M\Omega}$) were filled with a cesium-methylsulphonate ($CsMeSO_4$)-based internal pipette solution, containing 120 mM $CsMeSO_4$, 10 mM CsCl, 5 mM TEA-Cl, 1.5 mM $MgCl_2$, 10 mM HEPES, 0.1 mM EGTA, 2 mM Na-ATP, 0.5 mM Na-GTP and 5 mM QX-314, pH 7.25–7.30 with CsOH, 275–285 mOsm. The recording ACSF was supplemented with 5 µM GABA to replenish the extracellular GABA concentration reduced by the high-flow perfusion of the slices[32]. For recording I_{tonic} at $-70\,mV$, a high-CsCl-based internal solution was used, containing 140 mM CsCl, 1 mM $MgCl_2$, 10 mM HEPES, 0.1 mM EGTA, 4 mM NaCl, 2 mM Mg-ATP, 0.3 mM Na-GTP and 5 mM QX-314, pH ~7.3, ~$275\,\mathrm{mOsm\,l^{-1}}$, with ACSF containing 3 mM kynurenic acid to block glutamatergic currents.

Neurons were voltage clamped in whole-cell configuration using a MultiClamp-700A amplifier (Molecular Devices); all recordings were low-pass-filtered at 3 kHz (8-pole Bessel) and digitized online at 10 kHz (National Instruments PCI-MIO-16E-4 board). Series resistance and whole-cell capacitance were estimated from fast transients evoked by a 5-mV step and compensated to 75%. Excitatory postsynaptic currents (EPSCs) and inhibitory postsynaptic currents (IPSCs) were recorded by voltage clamping sequentially at $-70\,mV$ and $+10\,mV$, respectively.

All drugs were purchased from Sigma or Tocris. L-655,708 and SNAP-5114 were dissolved in DMSO then diluted 1:1,000 in H_2O. NO-711, gabazine and GABA were dissolved in H_2O.

Tonic inhibitory current and mean phasic current determination. Custom-written macros running under IGOR Pro v.6.0 (WaveMetrics) were used to analyse the digitized recordings to determine the values of tonic currents and mean phasic currents, as previously described[31]. I_{tonic} was recorded as the reduction in baseline holding currents (I_{hold}) after bath-applying a saturating amount ($>100\,µM$) of the $GABA_A$-receptor antagonist SR-95531 (gabazine), while voltage clamping at $+10\,mV$. NO-711, SNAP-5114 and L-655,708 were added to the recording ACSF via perfusion and their effects on I_{tonic} were recorded as the post-drug shift in I_{hold}. Drug perfusion was continued until the shifting I_{hold} remained steady for 1–2 min.

To determine the mean phasic current (I_{mean}), a 60-s segment containing either EPSCs or IPSCs was selected, and an all-point histogram was plotted for every 10,000 points (every 1 s), smoothed, and fitted with a Gaussian to obtain the mean baseline current. All baseline mean values were then plotted and linear trends subtracted to normalize the mean baseline current to 0 pA. After baseline normalization, the values

of each 10,000 points (each 1 s) were averaged to yield the value of I_{mean} (in $\mathrm{pA\,s^{-1}}$) for each 1-s epoch. The averaged I_{mean} value of a 60-s segment was reported as the phasic I_{mean} value for either the spontaneous EPSC or IPSC. Synaptic event kinetics (that is, frequency, peak amplitude, 10–90% rise time and weighted decay time constant) are analysed by custom-written LabView-based software (EVAN), as previously described[32]. For comparison of the IPSC peak amplitudes under control and PTX-treated conditions (Supplementary Table 2), the largest-amplitude count-matched method[32] was used, whereby the amplitude values in a given recording were sorted and the largest x number of events under control conditions were averaged and compared to the average of an equally matched x number of events recorded during a similar time period under the PTX condition, with x being the total number of events detected under the 10 µM PTX condition. This method circumvents the erroneous comparison of average amplitudes when considering the effects of a receptor antagonist that reduces the smaller events (in control conditions) down to the noise level.

Measurements of neuronal resting membrane potential and GABA reversal potential. To estimate neuronal resting membrane potential (V_{rest}), the cell-attached recording technique[35,36] was used. Briefly, depolarizing voltage ramps (-100 to $+200\,mV$) were applied to cell-attached patches to activate voltage-gated K^+ channels and establish the K^+ current reversal potential, which provides a measure of the V_{rest}, given near equimolar K^+ inside the cell and the pipette. The GABA reversal potential (E_{GABA}) was estimated by measuring the K^+ reversal potential after activating $GABA_A$ receptors with 50 µM muscimol. Recordings were made using a solution containing the following: 135 mM K^+ gluconate, 5 mM KCl, 2 mM $MgCl_2$, 10 mM HEPES, 0.1 mM EGTA, 4 mM Na-ATP, 0.3 mM Na-GTP, pH 7.3, $273\,\mathrm{mOsm\,l^{-1}}$. A junction potential of 9 mV was measured and then subtracted from voltage values of all measurements.

Fitting of multiple distributions to cumulative probability plots. The fitting of multiple distributions to a cumulative probability plot (Supplementary Fig. 2) was done as follows. Cumulative probabilities of the variable x (that is, $P(x)$) were calculated and fitted by one or more normal curves approximated by the logistic equation[37]:

$$P(x) = \sum_{i=1}^{n} R_i \frac{x^{p_i}}{x^{p_i} + \bar{x}_i^{p_i}}$$

where $R_1 \ldots R_n$ are the ratios of the n normal distributions (such as $\sum_{i=1}^{n} R_n = 1$), $\bar{x}_1 \ldots \bar{x}_n$ are the individual means, and $p_1 \ldots p_n$ are steepness factors related to the n standard deviations ($\mathrm{s.d.}_1 \ldots \mathrm{s.d.}_n$).

Grid-walking task. The grid-walking apparatus was manufactured as previously described[38,39], using a 12-mm square wire mesh with a grid area of 32 cm/20 cm/50 cm (length/width/height). A mirror was placed beneath the apparatus to allow video footage in order to assess the animals' stepping errors (foot faults). Each mouse was placed individually on top of the elevated wire grid and allowed to freely walk for a period of 5 min. Video footage was analysed offline by raters blind to the treatment groups. The total number of foot faults for each limb, along with the total number of non-foot-fault steps, was counted, and a ratio between foot faults and total steps taken calculated. Per cent foot faults were calculated by: number of foot faults / (foot faults + number of non-foot-fault steps) × 100. A ratio between foot faults and total steps taken was used to take into account differences in the degree of locomotion between animals and trials. A step was considered a foot fault if it was not providing support and the foot went through the grid hole. Furthermore, if an animal was resting with the grid at the level of the wrist, this was also considered a fault. If the grid was anywhere forward of the wrist area then this was considered as a normal step.

Spontaneous forelimb task (cylinder task). The spontaneous forelimb task encourages the use of forelimbs for vertical wall exploration/press in a cylinder[38,39]. When placed in a cylinder, the animal rears to a standing position, while supporting its weight with either one or both of its forelimbs on the side of the cylinder wall. Mice were placed inside a Plexiglas cylinder (15 cm in height with a diameter of 10 cm) and videotaped for 5 min. Videotaped footage of animals in the cylinder was evaluated quantitatively in order to determine forelimb preference during vertical exploratory movements. While the video footage was played in slow motion (1/5th real time speed), the time (seconds) during each that each animal spent on either the right forelimb, the left forelimb, or on both forelimbs were calculated. Only rears in which both forelimbs could be clearly seen were timed. The percentage of time spent on each limb was calculated and these data were used to derive an SFL asymmetry index ((per cent ipsilateral use) $-$ (per cent contralateral use)). The 'contact time' method of examining the behaviour was chosen over the 'contact placement' method, as described in ref. 38, as it takes into account the slips that often occur during a bilateral wall press after photothrombosis.

Western blot. Seven days after stroke mice were decapitated, the brains rapidly removed and peri-infarct cortex microdissected and frozen ($n = 5$). The equivalent region of cortex was taken in control, non-operated mice ($n = 3$). Samples were homogenized in radioimmunoprecipitation (RIPA) buffer (Pierce) and centrifuged at $20,000g$ at $4\,°C$ for 10 min. Supernatant was collected as protein extract

and stored at −80 °C. Western blot was performed as described[40]. One-hundred micrograms of protein from each sample was diluted in 7.5 μl of 2× SDS-sample buffer gel (Invitrogen) containing dithiothreitol (DTT) (Sigma) and brought to a final volume of 15 μl with RIPA buffer. Samples were denatured at 95 °C, loaded on to a 4–12% gradient Tris-Glycine gel (Invitrogen), separated via SDS–polyacrylamide gel electrophoresis (PAGE), and then transferred to HYBOND-P (PVDF) membrane (Amersham) at 30 V for 2 h. Membranes were rinsed and blocked overnight at 4 °C. Membranes were probed with antibodies against GABA transporter 3 (rabbit anti-GAT-3 1:1,000; Millipore), and GABA transporter 1 (rabbit anti-GAT-1 1:200; Millipore). Following successive washes, membranes were incubated in IgG donkey anti-rabbit horseradish peroxidise (HRP)-labelled secondary (1:6,000; Jackson) for 1 h at room temperature (21–23 °C). Membranes were incubated in ECL PLUS (Amersham) and chemiluminescence was detected using Fluorochem (Alpha Innotech). Membranes were then re-probed for 1 h at room temperature (21–23 °C) with GAPDH (1:2,500; Abcam) and donkey anti-rabbit-HRP (1:10,000; Jackson) as an endogenous control protein to ensure equal loading. Immunoblotting was performed in triplicate for each antibody. Adobe Photoshop software (Adobe Systems) was used for densitometric analysis of all blots.

31. Lee, K., Kim, J. E., Sivula, M. & Strittmater, S. M. Nogo receptor antagonism promotes stroke recovery by enhamcing axonal plasticity. *J. Neurosci.* **24,** 6209–6217 (2004).

32. Glykys, J., Mann, E. O. & Mody, I. Which GABA_A receptor subunits are necessary for tonic inhibition in the hippocampus? *J. Neurosci.* **28,** 1421–1426 (2008).

33. STAIR. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke.* **30,** 2752–2758 (1999).

34. Fisher, M., Feuerstein, G., Howells, D. W., Hurn, P. D., Kent, T. A., Savitz, S. I. & Lo, E. H. Update of the stroke therapy academic industry roundtable preclinical recommendations. *Stroke.* **40,** 2244–2250 (2009).

35. Stell, B. & Mody, I. Receptors with different affinities mediate phasic and tonic GABA_A conductances in hippocampal neurons. *J. Neurosci.* **22,** RC223 (2002).

36. Verheugen, J. A., Fricker, D. & Miles, R. Noninvasive measurements of the membrane potential and GABAergic action in hippocampal interneurons. *J. Neurosci.* **19,** 2546–2555 (1999).

37. Barlow, R. Cumulative frequency curves in population analysis. *Trends Pharmacol. Sci.* **11,** 404–406 (1990).

38. Baskin, Y. K., Dietrich, W. D. & Green, E. J. Two effective behavioral tasks for evaluating sensorimotor dysfunction following traumatic brain injury in mice. *J. Neurosci. Methods* **129,** 87–93 (2003).

39. Schallert, T., Fleming, S. M., Leasure, J. L., Tillerson, J. L. & Bland, S. T. CNS plasticity and assessment of forelimb sensorimotor outcome in unilateral rat models of stroke, cortical ablation, parkinsonism and spinal cord injury. *Neuropharmacology* **39,** 777–787 (2000).

40. Moore, C. S. *et al.* Increased X-linked inhibitor of apoptosis protein (XIAP) expression exacerbates experimental autoimmune encephalomyelitis (EAE). *J. Neuroimmunol.* **203,** 79–93 (2008).

Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration

Bi-Sen Ding[1], Daniel J. Nolan[1], Jason M. Butler[1], Daylon James[1], Alexander O. Babazadeh[1], Zev Rosenwaks[2], Vivek Mittal[3], Hideki Kobayashi[1], Koji Shido[1], David Lyden[4], Thomas N. Sato[5], Sina Y. Rabbany[1,6] & Shahin Rafii[1]

During embryogenesis, endothelial cells induce organogenesis before the development of circulation[1–4]. These findings suggest that endothelial cells not only form passive conduits to deliver nutrients and oxygen, but also establish an instructive vascular niche, which through elaboration of paracrine trophogens stimulates organ regeneration, in a manner similar to endothelial-cell-derived angiocrine factors that support haematopoiesis[5–7]. However, the precise mechanism by which tissue-specific subsets of endothelial cells promote organogenesis in adults is unknown. Here we demonstrate that liver sinusoidal endothelial cells (LSECs) constitute a unique population of phenotypically and functionally defined VEGFR3$^+$CD34$^-$VEGFR2$^+$VE-cadherin$^+$FactorVIII$^+$CD45$^-$ endothelial cells, which through the release of angiocrine trophogens initiate and sustain liver regeneration induced by 70% partial hepatectomy. After partial hepatectomy, residual liver vasculature remains intact without experiencing hypoxia or structural damage, which allows study of physiological liver regeneration. Using this model, we show that inducible genetic ablation of vascular endothelial growth factor (VEGF)-A receptor-2 (VEGFR2) in the LSECs impairs the initial burst of hepatocyte proliferation (days 1–3 after partial hepatectomy) and subsequent reconstitution of the hepatovascular mass (days 4–8 after partial hepatectomy) by inhibiting upregulation of the endothelial-cell-specific transcription factor Id1. Accordingly, Id1-deficient mice also manifest defects throughout liver regeneration, owing to diminished expression of LSEC-derived angiocrine factors, including hepatocyte growth factor (HGF) and Wnt2. Notably, in in vitro co-cultures, VEGFR2-Id1 activation in LSECs stimulates hepatocyte proliferation. Indeed, intrasplenic transplantation of Id1$^{+/+}$ or Id1$^{-/-}$ LSECs transduced with Wnt2 and HGF (Id1$^{-/-}$Wnt2$^+$HGF$^+$ LSECs) re-establishes an inductive vascular niche in the liver sinusoids of the Id1$^{-/-}$ mice, initiating and restoring hepatovascular regeneration. Therefore, in the early phases of physiological liver regeneration, VEGFR2-Id1-mediated inductive angiogenesis in LSECs through release of angiocrine factors Wnt2 and HGF provokes hepatic proliferation. Subsequently, VEGFR2-Id1-dependent proliferative angiogenesis reconstitutes liver mass. Therapeutic co-transplantation of inductive VEGFR2$^+$Id1$^+$Wnt2$^+$HGF$^+$ LSECs with hepatocytes provides an effective strategy to achieve durable liver regeneration.

Sinusoidal endothelial cells (SECs) compose a structurally and functionally unique capillary network that vascularizes specific organs, including bone marrow and liver. In adult mice, bone marrow SECs, through expression of specific angiocrine trophogens, such as Notch ligands, support haematopoietic regeneration[5–7]. Similarly, the hepatic circulation is predominantly lined by LSECs[8–10], with each hepatocyte residing in cellular proximity to LSECs. However, the lack of phenotypic and operational definition of liver endothelial cells and paucity of

relevant mouse angiogenic genetic models[11–13] have handicapped studies of the role of LSECs in regulation of hepatic regeneration[14–18].

Here, we use a physiologically relevant partial hepatectomy model to elucidate the instructive role of LSECs in mediating hepatic regeneration (Supplementary Fig. 1). In contrast to the administration of hepatotoxic chemicals, which impairs the organization of LSECs and causes tissue hypoxia, cell death and inflammation (Supplementary Fig. 2)[8,13,19], in the partial hepatectomy model, resection of 70% of the liver mass without perturbing the integrity of the residual liver vasculature[11] activates hepatocyte regeneration[15–17]. As such, this model provides an instructive model for interrogating the role of structurally and functionally intact LSECs in supporting liver regeneration.

As the VEGF family plays a critical role in the regeneration of the bone marrow SECs[6], we hypothesized that VEGF receptors[20–22], including VEGFR2 or VEGFR3, also modulate LSEC function. Using VEGFR2-GFP mice in which the expression of green fluorescent protein (GFP) is driven by the native promoter of VEGFR2, we demonstrate that VEGFR2 and VEGFR3 are exclusively expressed in the liver endothelial cells but not in other liver cell types, including hepatocyte nuclear factor 4α (HNF4A)$^+$ hepatocytes (Fig. 1a and Supplementary Fig. 3). Notably, distribution of VEGFR3 expression is restricted to VEGFR2$^+$ LSECs that branch out from CD34$^+$VEGFR3$^-$ large vessels (Fig. 1b). Polyvariate flow cytometric analysis on non-parenchymal cells (NPCs) demonstrates the expression of endothelial-specific marker VE-cadherin on non-haematopoietic VEGFR3$^+$ VEGFR2$^+$CD45$^-$ LSECs, 97.6% of which are non-lymphatic (Prox1$^-$CD34$^-$)[22] endothelial cells expressing coagulation factor VIII (Fig. 1c, d). Thus we have designated a unique phenotypic and operational signature for LSECs of adult mice as VEGFR3$^+$ CD34$^-$VEGFR2$^+$VE-cadherin$^+$FactorVIII$^+$Prox-1$^-$CD45$^-$ vessels, distinguishing them from VEGFR3$^-$CD34$^+$VEGFR2$^+$VE-cadherin$^+$ CD45$^-$ non-sinusoidal endothelial cells and VEGFR3$^+$CD34$^+$Prox-1$^+$FactorVIII$^-$CD45$^-$ lymphatic endothelial cells. Identification of LSECs as VEGFR3$^+$CD34$^-$ and non-sinusoidal endothelial cells as VEGFR3$^-$CD34$^+$ is sufficient for quantification, purification and molecular profiling of LSECs.

To determine the mechanism by which LSECs regulate hepatic proliferation, we studied the regenerative kinetics of hepatocytes and LSECs after partial hepatectomy. Two days after partial hepatectomy, staining with VE-cadherin, hepatocyte marker epithelial (E)-cadherin and mitotic marker phosphorylated-histone-3 (P-H3) revealed that P-H3$^+$E-cadherin$^+$ mitotic hepatocytes were positioned in the proximity of non-proliferating LSECs (Fig. 1e). However, proliferation of LSECs starts at day 4 and plateaus by day 8 after partial hepatectomy (Fig. 1f and Supplementary Fig. 4). In comparison, quantification of P-H3$^+$HNF4A$^+$ hepatocytes showed that the rate of hepatocyte proliferation peaks during the first 4 days, levelling off by day 8 (Fig. 1g). These results suggest a chronologically biphasic contribution of LSECs

[1]Howard Hughes Medical Institute, Ansary Stem Cell Institute, and Department of Genetic Medicine, Weill Cornell Medical College, New York, New York 10065, USA. [2]Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine, New York, New York 10065, USA. [3]Department of Surgery, Weill Cornell Medical College, New York, New York 10065, USA. [4]Department of Pediatrics, Weill Cornell Medical College, New York, New York, 10065, USA. [5]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan. [6]Bioengineering Program, Hofstra University, Hempstead, New York 11549, USA.
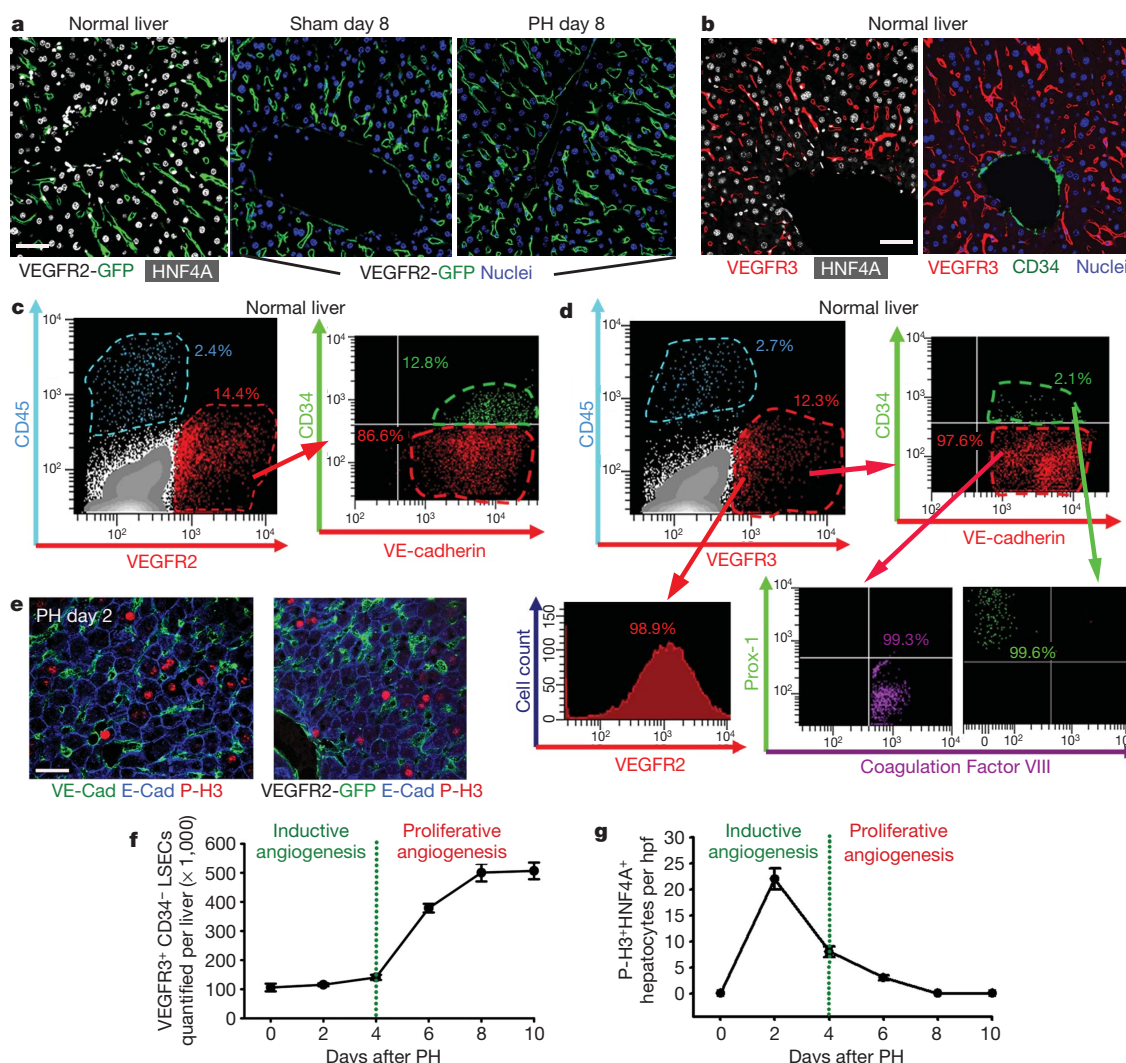
Figure 1 | Phenotypic signature and contribution of LSECs to physiological liver regeneration induced by 70% partial hepatectomy (PH). a, Liver sections obtained from VEGFR2-GFP reporter mice[6]. During liver regeneration VEGFR2 is exclusively expressed on the liver endothelial cells. **b**, Restricted expression of VEGFR3 on LSECs, but not CD34[+] large vessels or hepatocytes. **c**, Polyvariate flow cytometric analysis of the liver non-parenchymal cells. VEGFR2[+] cells that are CD45[−], express endothelial-specific VE-cadherin. **d**, Specific expression of VEGFR3 on VEGFR2[+]VE-cadherin[+]CD45[−] LSECs, with a predominant fraction being CD34[−]FactorVIII[+]Prox-1[−]. Thus LSECs could be identified as VEGFR3[+]CD34[−] cells. **e**, Forty-eight hours after partial hepatectomy, E-cadherin[+]P-H3[+] mitotic hepatocytes are localized adjacent to VE-cadherin[+] and VEGFR2[+] endothelial cells. **f, g**, Kinetics of LSEC expansion (**f**) and hepatocyte mitosis (**g**) during liver regeneration ($n = 4$); hpf, high-power field. Scale bars, 50 μm. Error bars, s.e.m.

in mediating hepatic reconstitution. At the early phases of partial hepatectomy (days 1–3 after partial hepatectomy), inductive angiogenesis in the non-proliferative LSECs stimulates hepatic regeneration, possibly by releasing angiocrine factors, whereas 4 days after partial hepatectomy, the increased demand of blood supply for the regenerating liver is met by proliferative angiogenesis of LSECs.

To investigate the significance of VEGF receptors during LSEC-driven hepatic regeneration, we designed experiments to delete the *VEGFR2* gene conditionally by crossing *VEGFR2[loxP/loxP]* mice with *ROSA-CreER[T2]* mice, generating inducible VEGFR2-deficient, *VEGFR2[flox/flox]* (*VEGFR2[fl/fl]*) mice (Supplementary Fig. 5)[6]. Owing to the endothelial-cell-specific expression of VEGFR2 in the liver, in *VEGFR2[fl/fl]* mice only liver endothelial cells, but not non-endothelial cells, will manifest functional defects. As control, we used mice with heterozygous deletion of the *VEGFR2* gene (*VEGFR2[fl/+]*). Forty-eight hours after partial hepatectomy, bromodeoxyuridine[+] hepatocyte proliferation (BrdU[+]HNF4A[+] cell number) was decreased by 67% in *VEGFR2[fl/fl]* mice (Fig. 2a, b). Notably, despite the patency of

the VE-cadherin[+]isolectin[+] perfused vessels at this early phase, the regeneration of liver mass was attenuated in *VEGFR2[fl/fl]* mice (Fig. 2c). Therefore, in the early phases (partial hepatectomy days 1–3) of the liver regeneration, targeting VEGFR2 primarily impairs the effect of endothelial-derived angiocrine factors to induce hepatocyte regeneration, but not vascular perfusion capacity.

However, in *VEGFR2[fl/fl]* mice at the later stages of liver regeneration (partial hepatectomy days 4–8), proliferative angiogenesis was also defective (Fig. 2c), interfering with the assembly of patent VE-cadherin[+]isolectin[+] vasculature (Fig. 2d, e), thereby blunting restoration of the liver mass for at least 28 days (Supplementary Fig. 5). Furthermore, in *VEGFR2[fl/fl]* mice, liver function after partial hepatectomy was abnormal, as manifested by elevated plasma bilirubin levels. To corroborate the endothelial-specific VEGFR2 function in mediating liver regeneration, *VEGFR2[loxP/loxP]* mice were also crossed with *VE-cadherin-CreER[T2]* mice to induce endothelial-selective deletion of VEGFR2 (Supplementary Fig. 5). Both the liver mass and formation of perfused vessels in the *VE-cadherin-CreER[T2]VEGFR2[fl/fl]*
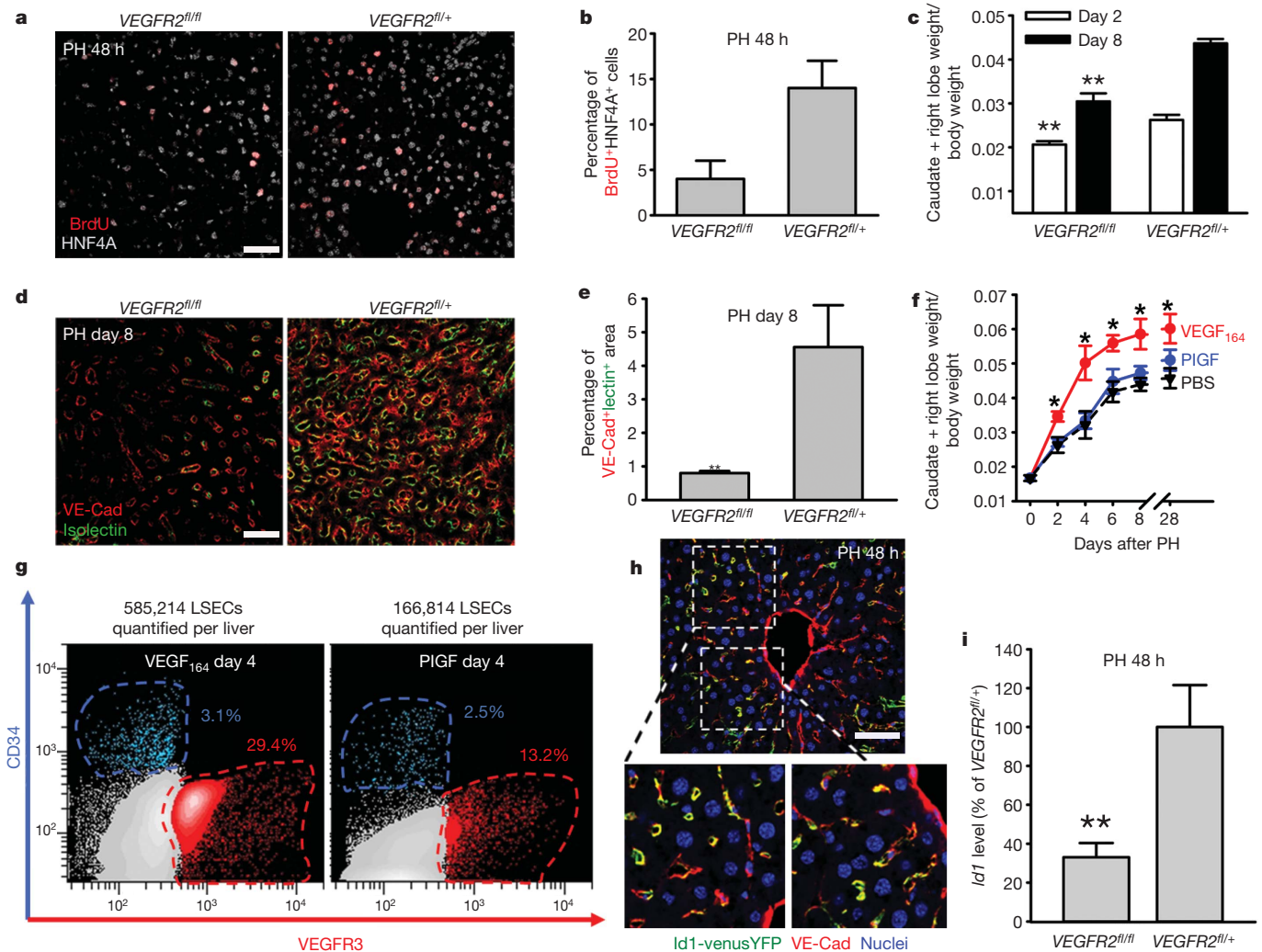
Figure 2 | VEGFR2-Id1 activation in LSECs mediates liver regeneration induced by partial hepatectomy. a, b, Hepatocyte proliferation after partial hepatectomy is impaired in $VEGFR2^{fl/fl}$ mice ($n = 5$). **c–e,** Inhibition of liver mass regeneration (**c**) and functional VE-cadherin[+]isolectin[+] vessel formation (**d, e**) in $VEGFR2^{fl/fl}$ mice after partial hepatectomy ($n = 4$–6). **f, g,** Injection of VEGF-A$_{164}$, but not VEGFR1-specific ligand PlGF, accelerates the regeneration of liver mass (**f**), associated with an incremental increase in VEGFR3[+]CD34[−] LSEC number (**g**) ($n = 4$). **h,** Regenerative liver section of $Id1^{VenusYFP}$ mouse[24]. $Id1$ is selectively upregulated by partial hepatectomy in VE-cadherin[+] vessels. **i,** VEGFR2 deletion diminishes $Id1$ upregulation in the regenerative liver ($n = 5$). *$P < 0.05$; **$P < 0.01$, versus $VEGFR2^{fl/+}$ (**b–e, i**), versus PlGF-treated group (**f**). Scale bar, 50 μm. Error bars, s.e.m.

mice were decreased after partial hepatectomy, which emphasizes the significance of VEGFR2 in mediating liver regeneration. Indeed, if the VEGF-A/VEGFR2 pathway promotes the LSEC-driven hepatic regeneration, then VEGF-A should enhance liver regeneration. Hence we compared the effect of VEGF-A$_{164}$ with placental growth factor (PlGF), as the latter selectively activates only VEGFR1[21]. After partial hepatectomy, VEGF$_{164}$, but not PlGF, accelerated the regeneration of both liver mass and the number of VEGFR3[+]CD34[−] LSECs, which were sustained for at least 28 days (Fig. 2f, g). Therefore, after partial hepatectomy, the activation of VEGF-A/VEGFR2, but not PlGF/VEGFR1, is crucial for priming LSECs to initiate and maintain hepatic proliferation.

To identify the angiocrine signals that stimulate liver regeneration, we used microarray analysis (Supplementary Fig. 6 and Supplementary Table 1). Among the endothelial-specific genes, the transcription factor $Id1$ was specifically upregulated in the endothelial cells activated by partial hepatectomy[23]. Using $Id1^{venusYFP}$ reporter mice in which the venusYFP expression is driven by the $Id1$ promoter[24], we found exclusive $Id1$ upregulation in LSECs 48 h after partial hepatectomy (Fig. 2h), which was significantly blunted in $VEGFR2^{fl/fl}$ mice (Fig. 2i). Remarkably, the liver mass recovery in $Id1$-deficient ($Id1^{−/−}$) mice after partial hepatectomy was impaired for 28 days and remained

unchanged upon VEGF-A$_{164}$ administration (Fig. 3a and Supplementary Fig. 7). Furthermore, after partial hepatectomy, $Id1^{−/−}$ mice exhibited significant decrease in mitotic BrdU[+]HNF4A[+] hepatocyte number, disrupted formation of functional VE-cadherin[+]isolectin[+] vessels, diminished proliferation of VEGFR3[+]CD34[−] LSECs, and abnormal liver function, as evidenced by an increase in plasma bilirubin levels (Fig. 3b, c and Supplementary Fig. 7). Thus activation of the VEGF-A/VEGFR2 pathway through upregulation of Id1 drives liver regeneration.

The role of Id1 upregulation in mediating the angiocrine function of LSECs on hepatocyte proliferation was also examined by an LSEC–hepatocyte co-culture system. Co-incubation of isolated hepatocytes with primary LSECs led to a ninefold increase in hepatocyte number, which was selectively abolished by knockdown of $Id1$ in LSECs (Fig. 3d, e and Supplementary Fig. 8). Conditioned medium from LSECs failed to support hepatocyte growth, underlining the importance of cell–cell contact in LSEC-derived angiocrine function. Therefore lack of $Id1$ results in defective inductive function of LSECs, impairing hepatocyte regeneration.

To determine whether *in vivo* angiocrine effects of $Id1^{+/+}$ LSECs could initiate hepatocyte regeneration in $Id1^{−/−}$ mice, we used the intrasplenic transplantation approach on day 2 after partial hepatectomy to engraft LSECs into the $Id1^{−/−}$ liver vasculature (Fig. 3f)[25].
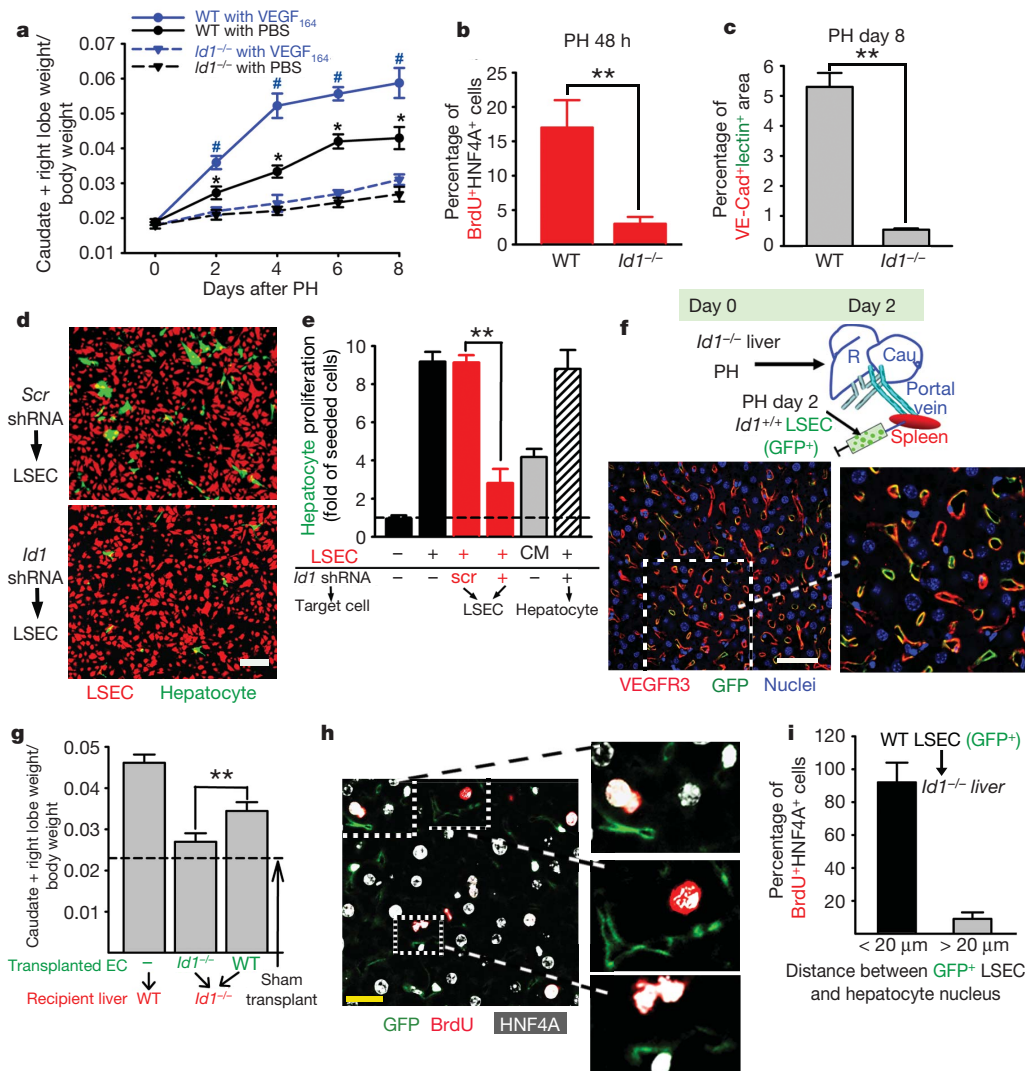
Figure 3 | Id1 upregulation in LSECs is essential for liver regeneration.
a, Compared with their wild-type (WT) littermates, $Id1^{-/-}$ mice manifest impaired regeneration in liver mass, which fails to be rescued by VEGF-A$_{164}$ administration ($n = 5$). **b, c,** Impaired hepatocyte proliferation (**b**) and assembly of VE-cadherin$^+$isolectin$^+$ vessels (**c**) in the $Id1^{-/-}$ mice after partial hepatectomy ($n = 5$). **d, e,** The LSEC-dependent stimulation of hepatocyte proliferation was specifically inhibited by $Id1$ gene knockdown. Scr, scrambled. CM, LSEC-conditioned medium ($n = 4$). **f,** Intrasplenic transplantation of GFP-marked LSECs incorporated into the lumen of VEGFR3$^+$ sinusoidal vasculature in the $Id1^{-/-}$ liver[25].**g, h,** Transplantation of $Id1^{+/+}$ LSECs restores the regeneration of mass (**g**) and hepatocyte proliferation (**h**) in the $Id1^{-/-}$ liver ($n = 4$). Dashed line, level of $Id1^{-/-}$ liver without endothelial cell transplantation. **i,** Cellular proximity is essential in the stimulation of hepatocyte mitosis by the transplanted $GFP^+Id1^{+/+}$ vasculature. *$P < 0.05$, versus $Id1^{-/-}$ (**a**); **$P < 0.01$, versus $Id1^{-/-}$ with VEGF$_{164}$ (**a**), versus WT (**b, c**). Scale bars, 50 μm (**d, f**) and 20 μm (**h**). Error bars, s.e.m.

GFP-marked $Id1^{+/+}$ LSECs selectively incorporated into the VEGFR3$^+$ sinusoidal vascular lumen and restored the regeneration of liver mass and LSEC expansion (Fig. 3g and Supplementary Fig. 9). In contrast, the transplanted $Id1^{-/-}$ LSECs failed to restore the regeneration of the $Id1^{-/-}$ liver. Moreover, in the $Id1^{-/-}$ liver, transplantation of $GFP^+ Id1^{+/+}$ LSECs at day 2 after partial hepatectomy initiated the proliferation of the hepatocytes in their immediate proximity (Fig. 3h, i). Thus partial vascular chimaerism afforded by the incorporation of $Id1$-competent LSECs generates sufficient endothelial-cell-derived inductive signals to initiate hepatic proliferation in the $Id1^{-/-}$ liver.

To identify endothelial-derived angiocrine factors that induce liver regeneration, we analysed LSECs purified from the wild-type and $Id1^{-/-}$ mice 48 h after partial hepatectomy. Among the known hepatic trophogens[10,18,26–28], the expression of Wnt2 and HGF, but not other trophogens expressed by LSECs, such as Wnt9B and thrombo-modulin, were drastically diminished in $Id1^{-/-}$ LSECs (Fig. 4a and Supplementary Fig. 10). These results suggest that Id1 upregulation in LSECs initiates hepatocyte proliferation through inducing Wnt2 and

HGF expression. To test this hypothesis, on day 2 after partial hepatectomy, we engrafted $Id1^{-/-}$ LSECs transduced with Wnt2, HGF or both into the $Id1^{-/-}$ liver vasculature by intrasplenic trans-plantation. Only $Id1^{-/-}$ LSECs carrying both Wnt2 and HGF ($Id1^{-/-} Wnt2^+ HGF^+$) restored the regeneration of mass and LSEC expansion in the $Id1^{-/-}$ liver (Fig. 4b), which suggests a collabora-tive effect between HGF and Wnt2. Notably, transplantation of $Id1^{-/-} Wnt2^+ HGF^+$ LSECs into $Id1^{-/-}$ mice increased the mitotic BrdU$^+$HNF4A$^+$ hepatocyte number to a similar degree achieved by $Id1^{+/+}$ LSEC transplantation (Fig. 4c). The mitotic hepatocytes were also found to be positioned adjacent to the transplanted $Id1^{-/-} Wnt2^+ HGF^+ GFP^+$ LSECs (Fig. 4d). Therefore Id1-activated LSECs through elaboration of Wnt2 and HGF induce proliferation of juxtaposed hepatocytes (Fig. 4e).

Here we have used conditional VEGFR2 knockout, $Id1^{-/-}$ mice, and an endothelial cell transplantation model to identify the essential angio-crine role of a specialized organ-specific vascular niche cell, defined operationally as VEGFR3$^+$CD34$^-$VEGFR2$^+$VE-cadherin$^+$FactorVIII$^+$ Prox1$^-$CD45$^-$ LSECs, in orchestrating physiological liver regeneration
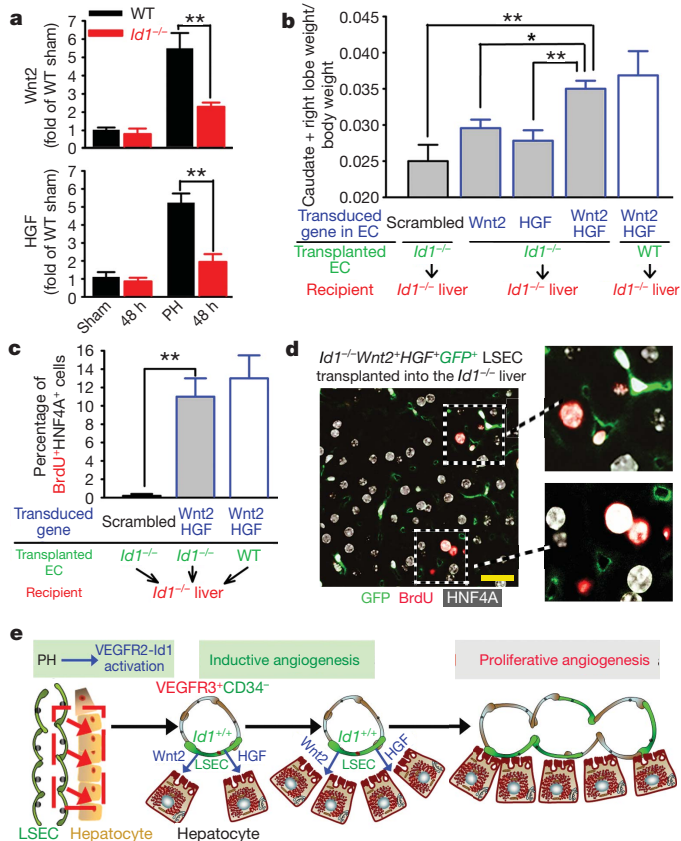
Figure 4 | Id1-mediated induction of Wnt2 and HGF in LSECs stimulates hepatic regeneration. a, Upregulation of HGF and Wnt2 is impaired in $Id1^{-/-}$ LSECs after partial hepatectomy ($n = 5$). **b**, Intrasplenic transplantation of GFP-marked $Id1^{-/-}$ LSECs carrying both Wnt2 and HGF ($Id1^{-/-}Wnt2^+HGF^+GFP^+$) rescues the regeneration of $Id1^{-/-}$ liver mass ($n = 4$). **c**, Transplantation of $Id1^{-/-}Wnt2^+HGF^+$ LSECs restores the impaired hepatocyte proliferation in the $Id1^{-/-}$ liver ($n = 4$). **d**, The proximity between the mitotic hepatocytes and the $Id1^{-/-}Wnt2^+HGF^+GFP^+$ LSECs in the $Id1^{-/-}$ liver. **e**, Requirement for VEGFR2-Id1 pathway in LSEC-mediated liver regeneration. Intrasplenic transplantation of $Id1^{+/+}$ LSECs into the $Id1^{-/-}$ liver sinusoids restores hepatic-vascular regeneration. Transplanted $Id1^{+/+}$ or $Id1^{-/-}Wnt2^+HGF^+GFP^+$ LSECs localize to the vicinity of hepatocytes, promoting inductive and proliferative angiogenesis thereby sustaining physiological liver regeneration. *$P < 0.05$; **$P < 0.01$. Scale bar, 20 μm. Error bars, s.e.m.

induced by partial hepatectomy. Similar to upregulation of Id1 in the angiogenic tumour vessels[23], Id1 expression is minimal in the normal LSECs, but after partial hepatectomy activation of VEGFR2 induces exclusive upregulation of Id1 in the angiogenic LSECs. We demonstrate that in the first 3 days after partial hepatectomy, activation of the VEGFR2-Id1 pathway switches on an inductive angiogenesis program in non-proliferative $VEGFR3^+CD34^-VEGFR2^+Id1^+$ LSECs, which, through production of angiocrine factors Wnt2 and HGF, provokes hepatic proliferation. Subsequently, as the regenerating liver demands additional blood supply, VEGFR2-Id1-mediated proliferative angiogenesis of LSECs reconstitutes hepatovascular mass. Therefore, we introduce the concept that LSECs support liver regeneration through a biphasic mechanism: at the early phase immediately after partial hepatectomy, inductive angiogenic LSECs promote organogenesis through release of angiocrine factors, whereas proliferative angiogenic LSECs vascularize and sustain the expanding liver mass.

We show that transplantation of the $Id1^{-/-}Wnt2^+HGF^+$ LSECs into $Id1^{-/-}$ mice initiates and restores liver regeneration. This finding, and the observation that hepatic proliferation is severely blunted in the $VEGFR2$ and $Id1$-deficient mice, suggest that LSECs are chartered with the responsibility of establishing an inductive vascular niche to initiate

hepatic proliferation by elaborating angiocrine factors. Because isolation of LSECs for therapeutic liver regeneration might encounter technical difficulties, endothelial progenitor cells (EPCs) derived from non-hepatic tissues may alternatively substitute for LSECs to initiate and restore liver regeneration[29]. Notably, $VEGFR2^+Id1^+$ EPCs could initiate angiogenesis through release of angiocrine factors rather than structurally incorporating into vessel wall[29]. As such, intrahepatic transplantation of EPCs will open up new avenues of cell therapy to promote liver regeneration.

In the partial hepatectomy model used in our study, the vascular integrity of the residual liver lobes is maintained with minimal inflammatory response (Supplementary Fig. 2), thereby establishing an ideal model to study endothelial-dependent liver regeneration. However, in chemical (CCl_4)-induced liver injury models, severe vascular damage and cell death might require the recruitment of other non-endothelial cells, including stellate cells[19] and pro-angiogenic haematopoietic cells, such as $CXCR4^+VEGFR1^+$ hemangiocytes[30], to support liver regeneration.

However, here is one unsolved enigma: how is removal of 70% of the liver sensed by the LSECs in the residual liver to ignite hepatic proliferation[14–17]? Conceivably, the mass of the liver is maintained through continuous release of as yet unrecognized inhibitory factors. Removal of the liver shifts the balance towards the predominance of vascular excitatory factors, which activate LSECs. Likewise, an increase in the mass of the liver 4 days after partial hepatectomy instigates the release of factors that stimulate sprouting angiogenesis in LSECs. Subsequently, recovery of the liver to its developmentally predetermined baseline mass might re-establish as yet unidentified inhibitory signals that terminate the regenerative process. The rapid regeneration of the liver after partial hepatectomy requires collective and global proliferation of many hepatocytes. Indeed, as each hepatocyte resides in close proximity to LSECs, this remarkably harmonious activation of hepatocytes is achieved by switching on an angiocrine-dependent regenerative program to induce proliferation of mature hepatocytes throughout the residual liver after partial hepatectomy. Whether angiocrine factors could also promote the propagation of liver progenitor cells[14], in addition to the mature hepatocytes, remains to be investigated.

In our study, Wnt2 and HGF represent the predominant liver-specific angiocrine factors driving hepatic regeneration. As direct cellular contact between LSECs and hepatocytes was essential for proliferation of hepatocytes, it is conceivable that other angiocrine factors might collaborate with Wnt2 and HGF to modulate liver regeneration. For instance, endothelial-specific extracellular matrix components, proteases, adhesion molecules and chemokines might also participate in hepatogenesis. Our *in vitro* endothelial cell–hepatocyte co-culture model and *in vivo* intrasplenic transplantation model provide ideal models to assess the role of these unknown angiocrine factors in modulating hepatic homeostasis during recovery from chemical or traumatic injury.

Accumulating evidence suggests that, in addition to LSECs, other organ-specific vascular niches play a seminal role in organ repair and tumorigenesis[5–7]. For example, stress-induced expression of Notch ligands by the bone marrow SECs has been shown to be essential for haematopoietic stem cell reconstitution[5]. Furthermore, elaboration of specific prototypical angiocrine factors, such as BMP2, nitric oxide, FGF2 and PDGF-β, by tumour vessels also directly provokes tumour progression and metastasis[7]. Collectively, these data suggest that tissue-specific expression of defined angiocrine factors may dictate heterogeneity of vasculature in regulating developmental and adult organogenesis.

So far, attempts at liver regeneration by hepatocyte transplantation have culminated in limited success[25]. Our study indicates that co-transplantation of hepatocytes or their progenitor cells[14] with $VEGFR2^+Id1^+$ LSECs or EPCs might allow the design of effective strategies to rescue hepatovascular function in patients inflicted with traumatic or infectious liver damage. Furthermore, the fact that

physiological liver regeneration is dependent on the proper inductive and proliferative functioning of the LSECs also calls for the assessment of the potential increased risks of anti-angiogenic therapy in clinical trials involving liver regeneration.

METHODS SUMMARY

Transgenic reporter, gene targeted animals and mouse surgery. C57BL/6J mice were obtained from Jackson Laboratories. *VEGFR2-GFP* mice were acquired from J. Rossant. *VE-cadherin-CreER^T2* mice were provided by L. Iruela-Arispa. Inducible VEGFR2 knockout (generated by T. N. Sato) and *Id1^−/−* mice were previously described[6,23]. *Id1^venusYFP* mice were obtained from R. Benezra[24]. Partial hepatectomy was performed by resecting three most anterior lobes. Hepatic engraftment of endothelial cells was adapted as previously reported[25]. All animal experiments were performed under the guidelines set by the Institutional Animal Care and Use Committee.

Image acquisition, image analysis, and flow cytometric analysis. Fluorescent images were captured on AxioVert LSM510 or 710 confocal microscope (Zeiss). For flow cytometry, antibodies were conjugated to Alexa Fluorescent dyes or Qdots (Invitrogen). Purified liver cells were analysed on LSRII-SORP (BD Biosciences). Doublets were excluded by FSC-W × FSC-H and SSC-W × SSC-H analysis, and single-stained channels were used for compensation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
2. Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564–567 (2001).
3. Sakaguchi, T. F., Sadler, K. C., Crosnier, C. & Stainier, D. Y. Endothelial signals modulate hepatocyte apicobasal polarization in zebrafish. *Curr. Biol.* **18**, 1565–1571 (2008).
4. Makita, T., Sucov, H. M., Gariepy, C. E., Yanagisawa, M. & Ginty, D. D. Endothelins are vascular-derived axonal guidance cues for developing sympathetic neurons. *Nature* **452**, 759–763 (2008).
5. Butler, J. N. *et al.* Endothelial cells are essential for the self-renewal and repopulation of notch-dependent hematopoietic stem cells. *Cell Stem Cell* **6**, 1–14 (2010).
6. Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4**, 263–274 (2009).
7. Butler, J. M., Kobayashi, H. & Rafii, S. Instructive role of the vascular niche in promoting tumour growth and tissue repair by angiocrine factors. *Nature Rev. Cancer* **10**, 138–146 (2010).
8. McDonald, B. *et al.* Interaction of CD44 and hyaluronan is the dominant mechanism for neutrophil sequestration in inflamed liver sinusoids. *J. Exp. Med.* **205**, 915–927 (2008).
9. Lee, J. S., Semela, D., Iredale, J. & Shah, V. H. Sinusoidal remodeling and angiogenesis: a new function for the liver-specific pericyte? *Hepatology* **45**, 817–825 (2007).
10. Klein, D. *et al.* Wnt2 acts as a cell type-specific, autocrine growth factor in rat hepatic sinusoidal endothelial cells cross-stimulating the VEGF pathway. *Hepatology* **47**, 1018–1031 (2008).
11. Greene, A. K. *et al.* Endothelial-directed hepatic regeneration after partial hepatectomy. *Ann. Surg.* **237**, 530–535 (2003).
12. Van Buren, G., II *et al.* Effect of molecular therapeutics on liver regeneration in a murine model. *J. Clin. Oncol.* **26**, 1836–1842 (2008).
13. LeCouter, J. *et al.* Angiogenesis-independent endothelial protection of liver: role of VEGFR-1. *Science* **299**, 890–893 (2003).
14. Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science* **322**, 1490–1494 (2008).
15. Fausto, N., Campbell, J. S. & Riehle, K. J. Liver regeneration. *Hepatology* **43**, S45–S53 (2006).
16. Michalopoulos, G. K. & DeFrances, M. C. Liver regeneration. *Science* **276**, 60–66 (1997).
17. Greenbaum, L. E. *et al.* CCAAT enhancer- binding protein beta is required for normal hepatocyte proliferation in mice after partial hepatectomy. *J. Clin. Invest.* **102**, 996–1007 (1998).
18. Huh, C. G. *et al.* Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc. Natl Acad. Sci. USA* **101**, 4477–4482 (2004).
19. Friedman, S. L. Hepatic stellate cells: protean, multifunctional, and enigmatic cells of the liver. *Physiol. Rev.* **88**, 125–172 (2008).
20. Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
21. Carmeliet, P. Angiogenesis in life, disease and medicine. *Nature* **438**, 932–936 (2005).
22. Alitalo, K., Tammela, T. & Petrova, T. V. Lymphangiogenesis in development and human disease. *Nature* **438**, 946–953 (2005).
23. Lyden, D. *et al.* Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401**, 670–677 (1999).
24. Nam, H. S. & Benezra, R. High levels of Id1 expression define B1 type adult neural stem cells. *Cell Stem Cell* **5**, 515–526 (2009).
25. Follenzi, A. *et al.* Transplanted endothelial cells repopulate the liver endothelium and correct the phenotype of hemophilia A mice. *J. Clin. Invest.* **118**, 935–945 (2008).
26. Goessling, W. *et al.* Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. *Cell* **136**, 1136–1147 (2009).
27. Ober, E. A., Verkade, H., Field, H. A. & Stainier, D. Y. Mesodermal Wnt2b signalling positively regulates liver specification. *Nature* **442**, 688–691 (2006).
28. Thompson, M. D. & Monga, S. P. WNT/beta-catenin signaling in liver health and disease. *Hepatology* **45**, 1298–1305 (2007).
29. Rafii, S. & Lyden, D. Cancer. A few to flip the angiogenic switch. *Science* **319**, 163–164 (2008).
30. Jin, D. K. *et al.* Cytokine-mediated deployment of SDF-1 induces revascularization through recruitment of CXCR4^+ hemangiocytes. *Nature Med.* **12**, 557–567 (2006).

METHODS

Transgenic reporter and gene targeted animals. C57BL/6J mice were obtained from Jackson Laboratories.VEGFR2-GFP mice were acquired from J. Rossant[31]. $Id1^{-/-}$ mice were generated as previously described[24] and obtained from R. Benezra and D. Lyden.

$VEGFR2^{loxP/loxP}$ mouse was generated by T. N. Sato and experiments with endothelial-specific inducible VEGFR2 knockout mice were performed as previously described[6]. Briefly, the $VEGFR2^{loxP/loxP}$ mice were bred with $RosaCre-ER^{T2}$ transgenic mice to establish the $RosaCre-ER^{T2}VEGFR2^{loxP/loxP}$ line and control $ROSA-CreER^{T2}VEGFR2^{loxP/+}$ to account for potential Cre-mediated toxicity. To induce endothelial-specific knockdown of VEGFR2, $VE-cadherin-CreER^{T2}$ mice provided by L. Iruela-Arispa were also crossed with $VEGFR2^{loxP/loxP}$ mice to generate $VE-cadherin-CreER^{T2}VEGFR2^{loxP/loxP}$ mice. To induce VEGFR2 gene ablation, 6- to 8-week-old male mice were treated with tamoxifen at a dose of $250\,mg\,kg^{-1}$ sunflower oil intraperitoneally for 6 days, interrupted for 3 days after the third dose. After 3 days of respite, the fourth dose was reinstituted for an additional 3 days, resulting in $ROSA-CreER^{T2}VEGFR2^{flox/flox}$ ($VEGFR2^{fl/fl}$) mice that were deficient in VEGFR2 at both alleles, the control $ROSA-CreER^{T2}VEGFR2^{fl/+}$ mice or $VE-cadherin-CreER^{T2}VEGFR2^{fl/fl}$ mice that had endothelial-cell-specific VEGFR2 knockdown. All animal experiments were performed under the guidelines set by the Institutional Animal Care and Use Committee.

Mouse liver regeneration model. A 70% partial hepatectomy model was used to induce physiological liver regeneration in mice. Three most anterior lobes (right medial, left medial and left lateral lobes), which comprise 70% of the liver weight, were resected, without injuring the blood supply to the caudate and the right lobes. Mice were anaesthetized by $100\,mg\,kg^{-1}$ intraperitoneal ketamine and $10\,mg\,kg^{-1}$ xylazine. Midline laparotomy was performed in the anaesthetized mice. After opening the upper abdomen and the exposure of the liver, the left lobe to be resected was gently lifted while a 5-0 silk suture tie (Roboz) was placed underneath the lobe and positioned as proximal to the origin of the lobe as possible. The two ends of the suture were tied over the top of the liver lobe at the base of the lobe near the inferior vena cava. Three knots were tied, and microdissecting scissors was used to cut the tied lobe just distal to the suture. This process was repeated for the other median lobes to perform 70% partial hepatectomy. Then the peritoneum was re-approximated with a running 5-0 silk suture and the skin was closed with a running 4-0 silk suture.

Sham-operated mice underwent laparotomy without liver resection. To characterize the regeneration of liver mass and function, the weight of residual liver lobes were measured and normalized to mouse body weight at various time points days after partial hepatectomy, and plasma bilirubin levels were assayed (Genzyme Diagnostics) after 70% partial hepatectomy, respectively. To compare partial hepatectomy model to CCl_4-induced liver injury model, CCl_4 was intraperitoneally injected as previously described[13]. To test liver regeneration promoted by VEGF-A or PlGF, mice were treated with $15\,\mu g\,kg^{-1}$ of recombinant $VEGF_{164}$ (Biovision) and the same amounts of PlGF (Biovision) 12 h before the operation and twice a day thereafter. $Id1^{-/-}$ mice and wild-type littermates were also subjected to similar $VEGF_{164}$ and PBS treatment before and after operation.

Liver immunofluorescence and detection of GFP. VEGFR2-GFP, $VEGFR2^{fl/fl}$, $Id1^{-/-}$ and littermate control mice were subjected to partial hepatectomy or sham-operation, perfused with 4% paraformaldehyde, cryoprotected and snap frozen in OCT. For the analysis of the liver microvasculature, mice were intravenously injected with $2\,mg\,kg^{-1}$ *Griffonia simplicifolia* lectin (isolectin B4, Invitrogen) 5 min before being killed, as previously described[6]. For immunofluorescence microscopy, the liver sections (10 μm) were blocked (5% donkey serum/0.3% Triton X-100) and incubated in primary antibodies: anti-VEGFR3 monoclonal antibody (mAb, mF4-31C1, $10\,\mu g\,ml^{-1}$, ImClone), anti-VE-cadherin polyclonal Ab (pAb, $2\,\mu g\,ml^{-1}$, R&D Systems), anti-CD34 mAb (553731, $5\,\mu g\,ml^{-1}$, BD Biosciences), anti-phospho-Histone H3 (Millipore) and anti-HNF4A antibody (Abcam). After incubation in fluorophore-conjugated secondary antibodies ($2.5\,\mu g\,ml^{-1}$, Jackson ImmunoResearch), sections were counterstained with TOPRO3 or DAPI (Invitrogen).

Liver cell proliferation *in vivo* was measured by BrdU uptake. Briefly, mice received a single dose of BrdU (Sigma) intraperitoneally 60 min before death (at a dose of $50\,mg\,kg^{-1}$ animal weight). At the time of death, mice were anaesthetized, blood was collected from the inferior vena cava, and the remaining liver lobes were removed, weighed and further processed. Cryosections were stained using the BrdU Detection System (BD Biosciences) and fluorophore-conjugated secondary antibodies ($2.5\,\mu g\,ml^{-1}$, Jackson ImmunoResearch).

Image acquisition and image analysis. Immunohistochemistry images of liver sections were captured with AxioVision software (Zeiss) mounted on an Olympus BX51 microscope (Olympus America). Immunofluorescence images were captured on AxioVert LSM510 or 710 confocal microscope (Zeiss). Digital images were analysed for the density of endothelial marker (VE-cadherin$^+$) and functional perfused vessels (isolectin$^+$) using Image J (National Institutes of Health). Vessel density was expressed by the percentage of positive component to the total area in each high-power field, ×400.

Isolation and culture of mouse cells. Hepatocytes, LSECs, stellate and Kupffer cells were isolated from mice that underwent sham-operation and partial hepatectomy, by a two-step collagenase perfusion technique with modifications[32–36]. Briefly, after the inferior vena cava was cannulated and portal vein was cut, the liver was perfused at $5\,ml\,min^{-1}$ through the inferior vena cava with Liver Perfusion Medium (Invitrogen) at 37 °C for 10 min, followed by perfusion with Liver Digest Medium (Invitrogen) for an additional 10 min. The liver was dissociated in Hepatocyte Wash medium (Invitrogen), passed through dacron fabric with 70-μm pores and separated from the non-parenchymal hepatocyte depleted fraction (NPCs) by low-speed centrifugation ($50g \times 5\,min$), which were further purified by percoll gradient centrifugation, using stock Percoll solution as previously described[36]. The supernatant containing NPCs was collected and was washed twice at $50g$ for 5 min, pelleted at $350g$ for 7 min and fractionated with Percoll gradient centrifugation ($900g \times 20\,min$) with 75% stock Percoll solution and 35% stock Percoll solution, as previously described[36]. Fractions containing LSECs were enriched, mixed with an equal volume of PBS and centrifuged at $900g$ for 7 min. The pellet was washed with DMEM (Invitrogen) at $350g$ for 7 min and further labelled by mouse LSEC binding magnetic beads (Miltenyi). The purification of LSECs was performed according to the manufacturer's protocol. Purification of stellate and Kupffer cells was performed as previously described[33–35].

Flow cytometric analyses, identification and quantification of LSECs. Purified monoclonal antibodies were conjugated to Alexa Fluor dyes or Qdots per manufacturer's protocols (Molecular Probes/Invitrogen). Purified hepatocyte-depleted NPCs were analysed on LSRII-SORP (BD). Data were processed with FACSDiva 6.1 software (BD). Doublets were excluded by FSC-W × FSC-H and SSC-W × SSC-H analysis, single-stained channels were used for compensation and fluorophore minus one controls were used for gating. Monoclonal antibodies were purchased from BD except where noted: VE-cadherin (BV13, ImClone); VEGFR3 (mF4-31C1, ImClone); VEGFR2 (DC101, ImClone); CD45 (30-F11, BD Biosciences); CD34 (14-0341, eBioscience).

For quantification of LSECs, the livers were mechanically prepared as above and the number of SECs was quantified by co-staining with conjugated antibodies to VEGFR2, VEGFR3, VE-cadherin, CD34. The number of SECs equals the number of VEGFR3$^+$CD34$^-$VEGFR2$^+$VE-cadherin$^+$ cells. VEGFR3$^-$CD34$^+$VEGFR2$^+$VE-cadherin$^+$ cells were scored as non-SECs.

Determination of hepatocyte proliferation in co-culture with endothelial cells. Human LSECs were from ScienCell Research Laboratories. To knockdown $Id1$ selectively in LSECs, $Id1$/Scrambled short hairpin RNA (shRNA) lentiviruses were generated by co-transfecting 15 μg of shuttle lentiviral vector containing $Id1$/Scrambled shRNA, 3 μg of pENV/VSV-G, 5 μg of pRRE and 2.5 μg of pRSV-REV in 293T cells by Fugene 6 (Roche Applied Science). Viral supernatants were concentrated by ultracentrifugation. These concentrated viral preparations were used to transduce LSECs or hepatocytes. For co-culture studies, 10,000 isolated primary hepatocytes were plated in a 100-mm dish coated with type I collagen, seeded with 500,000 LSECs, or with LSECs treated with $Id1$/scramble shRNA lentivirus, respectively. Culture conditions consisted of Williams' E Medium (Invitrogen) supplemented with L-glutamine ($2\,mmol\,l^{-1}$), 1% fetal bovine serum (FBS), vascular endothelial growth factor-A ($VEGF-A_{164}$) ($5\,ng\,ml^{-1}$), dexamethasone at $10^{-9}\,mol\,l^{-1}$, streptomycin ($100\,U\,ml^{-1}$) and penicillin ($100\,U\,ml^{-1}$). Cells from each group were collected after 2 weeks. To visualize LSECs and hepatocytes, LSECs were marked by mCherry lentivirus (in pCCL backbone) as described above, and hepatocytes were infected with GFP lentivirus. Conditioned medium was also collected from 500,000 LSECs cultured for 2 weeks, filtered through a 0.22 μm filter, and added to 10,000 hepatocytes at 1:2 dilution, in the absence of LSEC co-culture. The numbers of LSECs and hepatocytes were assessed by flow cytometric analysis of mCherry and GFP signals. Hepatocyte proliferation was quantified by comparing the number of retrieved hepatocytes to the initially seeded hepatocyte number.

Affymetrix analysis and quantitative real-time PCR analysis. RNA was freshly isolated from the liver using RNeasy (Qiagen) and was converted to complementary DNA using Superscript II (Invitrogen). Microarray was performed using Mouse U133 2.0 (Affymetrix). Details of the methods for RNA quality, sample labelling, hybridization and expression analysis were according to the manual of the Affymetrix Microarray Kit. Quantitative PCR was performed using Taqman gene expression systems for mouse VEGFR2, VEGFR3, Id1, HGF, Wnt2, Wnt9B and TM (Applied Biosystems).

Liver transplantation of regenerative LSECs. Multi-lobular 70% partial hepatectomy was performed in wild-type ($Id1^{+/+}$) mice and age- and sex-matched $Id1^{-/-}$ mice. Forty-eight hours after partial hepatectomy, LSECs were isolated from wild-type mice ($Id1^{+/+}$ regenerative LSECs) and marked by GFP lentivirus (in pCCL backbone) transduction as described above. The transplantation

procedure was modified from that previously described[25]. Briefly, 48 h after partial hepatectomy, $Id1^{-/-}$ mice were anaesthetized and placed in the right lateral decubitus position. The left flank was scrubbed with Betadine, and the skin and abdominal wall were incised longitudinally (parallel to the spine). After the spleen was exteriorized, $Id1^{+/+}$ regenerative LSECs were injected into the parenchyma of the spleen through a 27-gauge needle. A splenectomy was performed after the injection. To compare the rescuing effect of $Id1^{+/+}$ regenerative LSECs, $Id1^{-/-}$ and wild-type mice also subjected to the intrasplenic injection of PBS and splenectomy 2 days after partial hepatectomy (sham transplant). To introduce Wnt2 and HGF expression in LSECs, Wnt2 and HGF complementary DNAs were purchased from Open Biosystems and cloned into lentiviral vector as described above. Infection of LSECs with virus encoding Wnt2 or HGF, or the same amounts of mixed Wnt2 and HGF, was performed with GFP lentivirus infection.

Data analysis. All data are presented as the mean $\pm$ s.e.m. of at least three separate experiments. Differences between groups were tested for statistical significance using Student's t-test or analysis of variance. Statistical significance was set at $P < 0.05$.

31. Ema, M., Takahashi, S. & Rossant, J. Deletion of the selection cassette, but not *cis*-acting elements, in targeted *Flk1-lacZ* allele reveals *Flk1* expression in multipotent mesodermal progenitors. *Blood* **107,** 111–117 (2006).
32. Tam, B. Y. *et al.* VEGF modulates erythropoiesis through regulation of adult hepatic erythropoietin synthesis. *Nature Med.* **12,** 793–800 (2006).
33. Passino, M. A., Adams, R. A., Sikorski, S. L. & Akassoglou, K. Regulation of hepatic stellate cell differentiation by the neurotrophin receptor p75NTR. *Science* **315,** 1853–1856 (2007).
34. Kumar, V. *et al.* Cell-derived anaphylatoxins as key mediators of antibody-dependent type II autoimmunity in mice. *J. Clin. Invest.* **116,** 512–520 (2006).
35. Winau, F. *et al.* Ito cells are liver-resident antigen-presenting cells for activating T cell responses. *Immunity* **26,** 117–129 (2007).
36. Kreamer, B. L. *et al.* Use of a low-speed, iso-density percoll centrifugation method to increase the viability of isolated rat hepatocyte preparations. *In Vitro Cell. Dev. Biol.* **22,** 201–211 (1986).

LETTER

A genome–wide RNAi screen reveals determinants of human embryonic stem cell identity

Na-Yu Chia[1,2]*, Yun-Shen Chan[1,3]*, Bo Feng[1]*, Xinyi Lu[1,4], Yuriy L. Orlov[5], Dimitri Moreau[6], Pankaj Kumar[6], Lin Yang[1], Jianming Jiang[1], Mei-Sheng Lau[1], Mikael Huss[5], Boon-Seng Soh[7], Petra Kraus[7], Pin Li[7], Thomas Lufkin[7], Bing Lim[7,8], Neil D. Clarke[5,9], Frederic Bard[6,9] & Huck-Hui Ng[1,2,3,4,9]

The derivation of human ES cells (hESCs) from human blastocysts represents one of the milestones in stem cell biology[1]. The full potential of hESCs in research and clinical applications requires a detailed understanding of the genetic network that governs the unique properties of hESCs. Here, we report a genome-wide RNA interference screen to identify genes which regulate self-renewal and pluripotency properties in hESCs. Interestingly, functionally distinct complexes involved in transcriptional regulation and chromatin remodelling are among the factors identified in the screen. To understand the roles of these potential regulators of hESCs, we studied transcription factor PRDM14 to gain new insights into its functional roles in the regulation of pluripotency. We showed that PRDM14 regulates directly the expression of key pluripotency gene POU5F1 through its proximal enhancer. Genome-wide location profiling experiments revealed that PRDM14 colocalized extensively with other key transcription factors such as OCT4, NANOG and SOX2, indicating that PRDM14 is integrated into the core transcriptional regulatory network. More importantly, in a gain-of-function assay, we showed that PRDM14 is able to enhance the efficiency of reprogramming of human fibroblasts in conjunction with OCT4, SOX2 and KLF4. Altogether, our study uncovers a wealth of novel hESC regulators wherein PRDM14 exemplifies a key transcription factor required for the maintenance of hESC identity and the reacquisition of pluripotency in human somatic cells.

Embryonic stem cells (ESCs) were first derived in mouse blastocysts as cell lines that can undergo extensive self-renewal *in vitro* and have the ability to undergo multi-lineage differentiation, also defined as pluripotency[2,3]. Although human ESCs (hESCs) share many similarities with mouse ESCs (mESCs), there are intriguing differences between these ESCs[4–6]. These differences could be due to species-specific differences in embryonic development. Alternatively, the ESCs could be derived from cells originating from different developmental stages. Consistent with this idea is the identification of post-implantation murine epiblast-derived stem cells which show characteristics of hESCs[7,8]. Hence, a comprehensive elucidation of regulators that specify the hESC identity is essential to understand the differences between pluripotent stem cells.

To define the genetic network that governs hESC identity, we carried out a whole genome RNA interference (RNAi) screen (Supplementary Fig. 1). A green fluorescent protein (GFP) reporter construct driven by the POU5F1 upstream regulatory region was introduced into H1 hESC to generate a stable hESC line and we used GFP fluorescence as a proxy for the undifferentiated state of hESCs (Fig. 1a and Supplementary Fig. 2). Using this reporter cell line, we screened a siRNA library targeting 21,121 human genes. The screen was carried out in duplicate and the cells were imaged for GFP and Hoechst fluorescence (Fig. 1a,

b). The mean of the z-score for GFP fluorescence reduction (F_{av}) and nuclei number reduction (N_{av}) were calculated. From the F_{av} candidate list (Supplementary Table 1), it is reassuring that POU5F1, which is essential for the maintenance of ESCs, was ranked first. In addition to that, we identified several regulators implicated in mESCs, including HCFC1, TCL1A, ZSCAN10, ZIC3, NANOG and ZNF143 to be among the top 5% of the gene list (Supplementary Fig. 3a). Genes that could affect survival and proliferation of hESCs could also be obtained by quantifying the nuclei number (Supplementary Table 2 and Supplementary Fig. 4).

We set a cutoff for genes with F_{av} z-score >2 to be considered potential candidates, and 566 genes were obtained (Fig. 1c). Gene ontology (GO) analysis of the 566 genes showed an enrichment for transcription factors and translation factors (Supplementary Fig. 3b). Reactome analysis also revealed enrichment of pathways involved in transcription and translation (Fig. 2a and Supplementary Table 3)[9]. Through the STRING database, we obtained protein–protein interactions for 263 out of the 566 genes (Supplementary Fig. 3c) and among them, we identified components of the INO80 chromatin remodelling complex[10], the mediator complex[11], the COP9 signalosome[12], the TAF complex[13], the eukaryotic initiation factor complex[14], and spliceosome complex[15] (Fig. 2b and Supplementary Fig. 3d, e). Hence, genes coding for proteins in known biochemical complexes which have not been previously implicated as important for hESCs were identified.

Next, we performed a secondary validation screen for 200 of the 566 candidates from the primary screen. The pooled siRNAs for each gene were deconvoluted into four individual siRNAs. Candidates were considered positive if they were scored by at least two siRNAs. To further enhance the confidence of the hit genes, we adopted a multiparametric approach where the importance of each gene in the maintenance of hESCs was assessed by different stemness markers (OCT4 and NANOG) for different hESCs cell lines (Supplementary Table 4). For POU5F1–GFP hESCs, the validation rate based on the reduction of GFP reporter, OCT4 and NANOG expression were 86.1%, 87.6% and 63.4%, respectively, and 127 genes were validated by all three markers. For HES2 hESCs, 86 common genes were obtained based on OCT4 and NANOG expression and the validation rate was 75.2% and 43.6%, respectively. Repeating with HES3 hESCs, we yielded 124 common genes with a validation rate of 64.9% and 72.3% for OCT4 and NANOG, respectively (Supplementary Fig. 5a, b and Supplementary Table 5). The higher validation rate for H1 POU5F1–GFP hESCs, compared to other hESC lines, corroborated the fact that this cell line was used for the primary screen. We identified 93 genes that downregulated OCT4 expression and 54 genes that downregulated NANOG expression in the three different hESC lines (Supplementary Fig. 5c and Supplementary Tables 6 and 7). In addition, we also observed a

Figure 1 | Genome-wide screen for regulators that maintain hESC identity.
a, Schematic representation of siRNA screen. H1 hESC line with a *GFP* reporter gene driven by the *POU5F1* promoter was used for the screen. For the primary screen, 21,121 genes were analysed. For secondary screen, individual siRNAs for 200 high-confidence candidates from the primary screen were selected for further validation using the reporter cells as well as two other hESC lines HES2 and HES3. **b**, Montage representing the images for GFP fluorescence (green) and Hoechst staining (blue) for a typical 384-well plate. The hESCs were imaged 4 days after transfection. The negative control siRNA (non-targeting) and the positive controls siRNA (*GFP* and *POU5F1*) were added to the designated wells as indicated. **c**, Dot plot of the genome-wide screen result. The *y*-axis represents the average *z*-scores for the GFP reduction for each targeted gene. Controls are represented by the black (non-targeting siRNA), red (*POU5F1* siRNA) and green (*GFP* siRNA) dots. Genes with *z*-score >2, highlighted in blue, are potential candidates required for the maintenance of hESC identity. The rest of the genes are indicated in orange.

positive correlation between the stemness markers (Supplementary Fig. 5d). We conducted further counter-screens using a hESC line harbouring a constitutively active *EF1a* (also known as EEF1A1)–*GFP* reporter and also assessed the level of control proteins β-actin and GAPDH (Supplementary Table 8). The results filtered out a minor fraction of the candidates that could be affecting global transcription or translation processes (Supplementary Fig. 6). Results from the secondary validation assay were used to select genes for further functional assays for pluripotency.

Pluripotency can be reinstated in somatic cells through the introduction of defined transcription factors such as OCT4, SOX2, KLF4 and c-Myc [16,17]. Hence, we used reprogramming as a gain-of-function assay to assess the role of PRDM14, NFRKB and YAP1 in the acquisition of pluripotency. PRDM14 is a transcriptional regulator with a potential enzymatic PR/SET domain whereas NFRKB is a component of the INO80 chromatin remodelling complex [10]. YAP1 served as a negative control. From the secondary validation screen, knockdown of *PRDM14* and *NFRKB*, but not *YAP1*, reduced OCT4 expression in the three hESCs. We transduced human fibroblasts with retroviruses containing expression constructs for OCT4 (O), SOX2 (S), KLF4 (K) and c-MYC (M) to generate human induced pluripotent stem cell (hiPSC) colonies that resembled hESCs. Interestingly, the number of hiPSC colonies was increased by 3.5-fold when we co-introduced PRDM14 (P) along with OSKM (Fig. 3a). NFRKB (N) enhanced reprogramming by twofold whereas YAP1 did not show a positive effect on reprogramming. Furthermore, knockdown of *PRDM14* or *NFRKB* affected the reprogramming process as fewer hiPSC colonies were obtained (Fig. 3b). The OSKMP and OSKMN hiPSCs showed similar characteristics and gene expression profile as hESCs (Supplementary Fig. 7). In the presence of PRDM14, an early induction of genes associated with initiation and maturation stages of reprogramming was observed, indicating that PRDM14 can accelerate these stages of the reprogramming process (Supplementary Fig. 8)[18]. Next, we showed that PRDM14 and NFRKB can enhance OSK-mediated reprogramming by 7- and 3.5-fold, respectively (Supplementary Fig. 9a–h). Interestingly, we also found that hiPSCs could be generated with OSMP and OSMN, indicating that PRDM14 and NFRKB can replace KLF4 (Supplementary Fig. 9i, j). However, no hiPSCs were obtained in the absence of OCT4 or SOX2. Taken together, we demonstrated that PRDM14 and NFRKB have important functions in the acquisition of pluripotency in human cells. This also highlights the value of the candidates that we identified in our screen.

The role of PRDM14 in regulating the pluripotency of hESCs was studied in greater detail as it was ranked highly in the primary screen. Furthermore, PRDM14 is highly expressed in a variety of hESCs [19] and is a target of the core transcription factors in hESCs [20]. PRDM14 has also been implicated to regulate self-renewal of hESCs since its depletion induced expression of differentiation marker genes and altered the cellular morphology [21]. We repeated the knockdown of *PRDM14* with four shRNA constructs in non-reporter hESCs. Depletion of PRDM14 resulted in a reduction in the expression of hESC-associated genes in three different hESC lines (H1, H9 and HES3) (Supplementary Figs 10a–c, 11 and 12). On the other hand, proteins that are expressed in differentiated cell-types were upregulated upon PRDM14 depletion (Supplementary Fig. 10d), indicating differentiation of hESCs. Importantly, the phenotypes induced by *PRDM14* shRNA could be rescued by co-expression of RNAi-immune *PRDM14* cDNA, indicating that the knockdown effect is specific (Supplementary Fig. 13).

In mice, *Prdm14* is essential for the establishment of the germ cell lineage, but not required for early embryonic development [22]. The knockdown of Prdm14 in mouse ESCs yielded no observable phenotypic change and no reduction in *Pou5f1* and *Sox2* (Supplementary Fig. 14a, b). Mouse epiblast-derived stem cells (EpiSCs) are also pluripotent and have features of hESCs [7,8], but they are deficient in *Prdm14* (Supplementary Fig. 14c, d). Thus, Prdm14 is required only for the maintenance of hESCs, but not mouse EpiSCs. It is noteworthy that other members of the Prdm family are expressed in mouse EpiSCs and these could be of functional importance (Supplementary Fig. 15).

To investigate further the function of PRDM14, we used chromatin immunoprecipitation coupled with sequencing (ChIP-seq) to map the PRDM14 binding loci (Supplementary Table 9). After validation of the ChIP-seq data set (Supplementary Fig. 16), we derived the consensus sequence motifs (Fig. 4a) by using the *de novo* motif-discovery algorithm, CisFinder [23]. Interestingly, PRDM14 colocalizes with OCT4, SOX2, NANOG and p300 (Fig. 4b). Co-motif analysis of the PRDM14
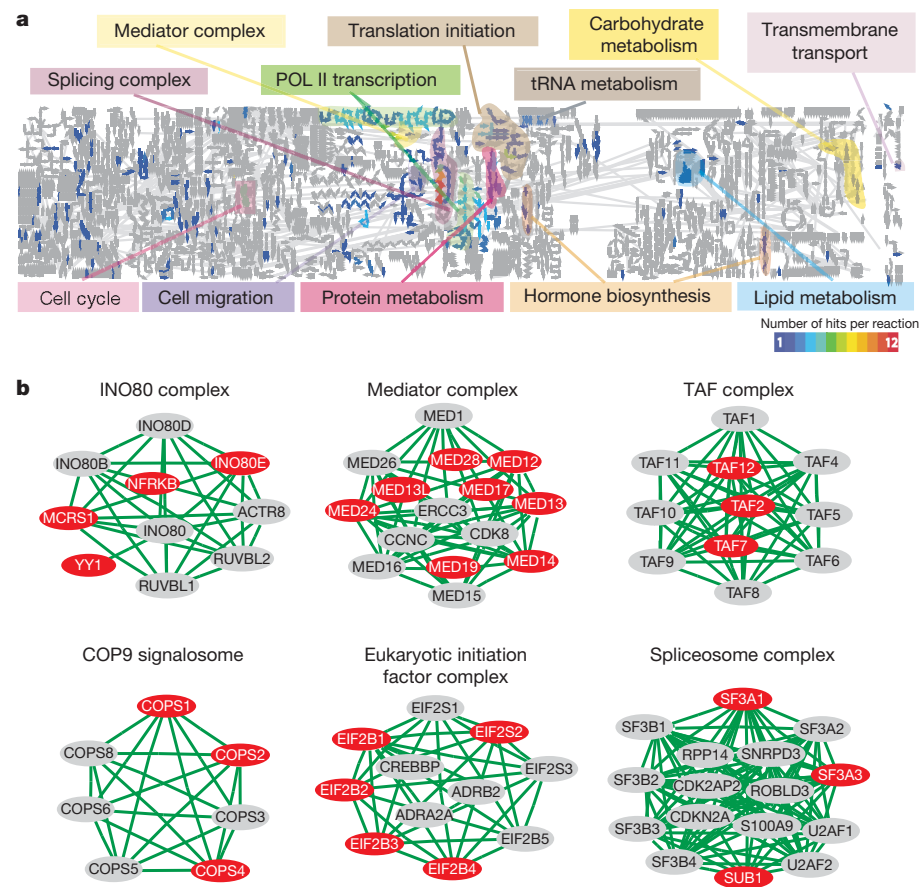
Figure 2 | Pathway analyses. a, Reactome analysis. The 566 genes (identifiers) were analysed using the web resource Reactome to determine the events (reactions and/or pathways) that were statistically over-represented. Twelve categories with P-value < 0.05 were over-represented. Single events were coloured according to the number of matching identifiers from blue (1 matching identifier) to red (12 matching identifiers). **b**, Components of the INO80 chromatin remodelling complex, mediator complex, TAF complex, COP9 signalosome, eukaryotic initiation complex and spliceosome complex with z-score > 2 are indicated in red.

binding loci showed a significant enrichment of a joint Sox2-Oct4 motif, confirming the co-occurrence of PRDM14, OCT4 and SOX2 sites (Supplementary Fig. 17). Our ChIP-seq analysis identified 2,755 genes that were bound by PRDM14 (Supplementary Table 10). Interestingly, we observed a PRDM14 peak at the *POU5F1* upstream regulatory region (Fig. 4c). This region is known to contain a proximal enhancer with conserved region 2 (CR2) and a distal enhancer with conserved

region 4 (CR4)[24–26]. Using electrophoretic mobility shift assay (EMSA), we confirmed that the CR2 sequence which contains the PRDM14 motif is indeed bound directly by both recombinant and native PRDM14 proteins (Supplementary Fig. 18a, b). Similar to the differential activity of the proximal and distal enhancers in mouse EpiSCs, the CR2 reporter is more active than the CR4 reporter in hESCs[7,8,24]. Depletion of PRDM14 led to a downregulation of CR2 enhancer activity (Supplementary Fig. 18c). Mutation of PRDM14 site at CR2 could also reduce its activity (Supplementary Fig. 18e). Using ChIP assay, we confirmed that PRDM14 binds to CR2 but not CR4 (Supplementary Fig. 18d). Furthermore, PRDM14 could activate a reporter construct containing the CR2 sequence in 293T cells (Fig. 4d). These data indicate that PRDM14 is directly activating the CR2 enhancer.

We further mapped the functional domains of PRDM14 with constructs expressing different fragments of the protein (Supplementary
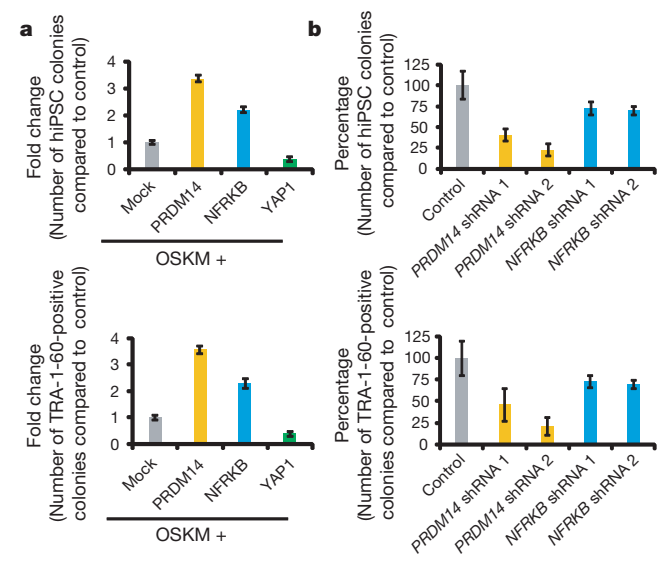
Figure 3 | PRDM14 and NFRKB can enhance reprogramming of human fibroblasts to iPSCs. a, Graph depicts fold change of the number of hiPSC colonies generated from PRDM14, NFRKB or YAP1 in conjunction with OCT4, SOX2, KLF4 and c-MYC (OSKM) with respect to the control (OSKM) (upper panel) and fold change of the number of TRA-1-60-positive colonies (lower panel). Each column represents the average of three replicates. **b**, PRDM14 is required for reprogramming of human somatic cell. Retroviruses harbouring *PRDM14* shRNA or *NFRKB* shRNA were cotransduced with the four reprogramming factors. Two independent shRNAs were used for the depletion of either *PRDM14* or *NFRKB*. The number of hESC-like iPSC (upper panel) or TRA-1-60-positive (lower panel) colonies was counted 4 weeks after infection. All values are means ± s.e.m. from three independent experiments ($n = 3$).
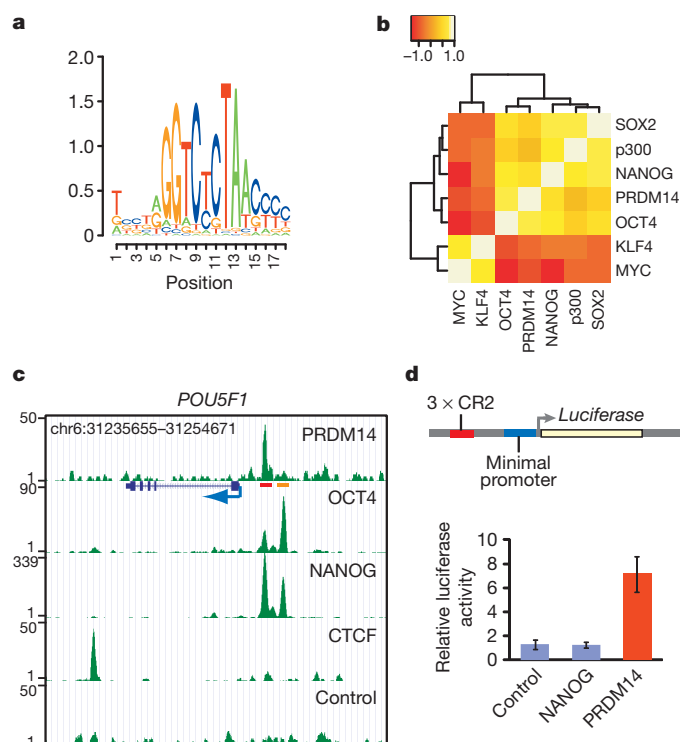
Figure 4 | PRDM14 is integrated into the core hESC transcriptional regulatory network. a, PRDM14 motif is predicted using the *de novo* motif-discovery algorithm CisFinder. **b**, PRDM14 shows co-binding with OCT4, SOX2, NANOG and co-activator p300. Colours in the heat map reflect the colocalization frequency of each transcription factor (the descending frequency of localization ranges from white to yellow and to red). **c**, ChIP-seq binding profile of PRDM14, OCT4, NANOG and CTCF at the *POU5F1* locus. Control ChIP-seq library was obtained from sequencing of input DNA. **d**, Three copies of CR2 (conserved region 2) consensus motif were inserted in tandem before the minimal promoter of the reporter construct. *NANOG* or *PRDM14* expression construct was cotransfected with the reporter construct into 293T cells and luciferase activity was normalized against the control vector. All values are means ± s.e.m. from three independent experiments ($n = 3$).

Fig. 19a). The putative DNA-binding domain and the amino-terminal region are required for transcriptional activation (Supplementary Fig. 19b). Importantly, we showed that these domains are required for the acquisition of pluripotency using the reprogramming assay (Supplementary Fig. 19c). The DNA binding activity resides within the carboxy-terminal zinc finger region as deleting this region abolished direct interaction with DNA (Supplementary Fig. 19d). Five of the six zinc fingers are required for DNA binding as well as transcriptional activity (Supplementary Fig. 19e, f).

The positive regulation of *POU5F1* expression by PRDM14 is unexpected as previous studies implicate PRDM14 as a transcriptional repressor[21,22]. To identify the genes which are regulated by PRDM14, we performed expression profiling upon PRDM14 depletion. We found that 358 of the 2,645 PRDM14-bound genes (13.5%) were downregulated and 638 of the PRDM14-bound genes (24.1%) were induced (Supplementary Figs 20a–d and 21 and Supplementary Tables 11 and 12). This finding indicates that PRDM14 can have both positive and negative roles on transcription. An OCT4 position weight matrix was enriched at the PRDM14 sites and PRDM14 ChIP-seq peaks were also observed to co-occur with OCT4, SOX2 and NANOG at PRDM14-regulated genes (Supplementary Fig. 22). Altogether, our genome-wide analyses showed that the target genes of PRDM14 are involved in diverse cellular processes (Supplementary Fig. 20e).

Identification of new molecules required to maintain the identity of ESCs is critical to the understanding of the mechanisms that govern self-renewal and pluripotency. Both small-scale and genome-wide

RNAi screens in murine ESCs[27–30] have uncovered many important regulators. Despite these efforts, little is known about the key players in hESCs. Given the substantial differences between the two, it is of interest to investigate the key genetic components that are specific to hESCs. Through a genome-wide RNAi screen, we identified novel factors that are essential for the maintenance of hESCs and characterized PRDM14 in detail as one of the hitherto unknown critical regulators of hESCs. Previous work on a Prdm14 knockout mouse model showed that Prdm14 is critical for the establishment of the germ cell lineage[22]. Hence, our study links a germ cell factor in mouse to the regulation of a key pluripotency gene in hESCs.

We revealed the involvement of several components of the different functionally distinct complexes (Fig. 2b) in the maintenance of hESCs. It is particularly intriguing to find that chromatin remodelling complex (INO80 complex), transcriptional regulatory complexes (mediator complex and TAF complex) and signalling complex (COP9 signalosome) are implicated in hESC biology. Future work will enable us to define more specifically their individual function and the mechanisms through which these regulators operate collectively in hESCs. Our study shows that these regulators may hold the potential in advancing the methodology and understanding of the mechanisms of human somatic cell reprogramming.

METHODS SUMMARY

POU5F1–GFP reporter hESC-line was generated for the genome-wide RNAi screen. The primary, secondary and counter screens were carried out in a 384-well plate format. Gene Ontology was performed with Panther classification (www.pantherdb.org). Pathway analyses were carried out using Reactome (www.reactome.org) and STRING (http://string.embl.de/). Individual shRNAs for each gene were designed using the WI siRNA selection program (http://jura.wi.mit.edu/bioc/siRNAext/). ChIP-seq data was generated using Illumina GA single-read sequencing and peak calling was carried out with MACS. For expression profiling experiments, RNA was amplified and hybridized to an Illumina microarray (HumanRef-8 v3.0 Expression BeadChips). pMXs retroviral plasmids that carry cDNAs of human *OCT4*, *SOX2*, *KLF4*, *c-MYC*, *PRDM14*, *NFRKB* and *YAP1* were used for reprogramming. Teratomas were generated 8 weeks after subcutaneous injection of hESCs or hiPSCs into SCID mice. Immunofluorescence staining, immunohistochemistry, ChIP, RT–PCR (PCR with reverse transcription), western blotting, co-immunoprecipitation and EMSA were performed using standard protocols.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–1147 (1998).
2. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
3. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78,** 7634–7638 (1981).
4. Yu, J. & Thomson, J. A. Pluripotent stem cell lines. *Genes Dev.* **22,** 1987–1997 (2008).
5. Van Hoof, D. *et al.* A quest for human and mouse embryonic stem cell-specific proteins. *Mol. Cell. Proteomics* **5,** 1261–1273 (2006).
6. Pera, M. F. & Trounson, A. O. Human embryonic stem cells: prospects for development. *Development* **131,** 5515–5525 (2004).
7. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).
8. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448,** 196–199 (2007).
9. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33,** D428–D432 (2005).
10. Conaway, R. C. & Conaway, J. W. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.* **34,** 71–77 (2009).
11. Casamassimi, A. & Napoli, C. Mediator complexes and eukaryotic transcription regulation: an overview. *Biochimie* **89,** 1439–1446 (2007).
12. Chamovitz, D. A. Revisiting the COP9 signalosome as a transcriptional regulator. *EMBO Rep.* **10,** 352–358 (2009).
13. Albright, S. R. & Tjian, R. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242,** 1–13 (2000).

14. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11,** 113–127 (2010).
15. Rino, J. & Carmo-Fonseca, M. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends Cell Biol.* **19,** 375–384 (2009).
16. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
17. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–146 (2008).
18. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
19. Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25,** 961–973 (2007).
20. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
21. Tsuneyoshi, N. *et al.* PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. *Biochem. Biophys. Res. Commun.* **367,** 899–905 (2008).
22. Yamaji, M. *et al.* Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature Genet.* **40,** 1016–1022 (2008).
23. Sharov, A. A. & Ko, M. S. Exhaustive search for over-represented DNA sequence motifs with CisFinder. *DNA Res.* **16,** 261–273 (2009).
24. Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4,** 513–524 (2009).
25. Nordhoff, V. *et al.* Comparative analysis of human, bovine, and murine *Oct-4* upstream promoter sequences. *Mamm. Genome* **12,** 309–317 (2001).
26. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122,** 881–894 (1996).
27. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442,** 533–538 (2006).
28. Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* **134,** 162–174 (2008).
29. Ding, L. *et al.* A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4,** 403–415 (2010).
30. Hu, G. *et al.* A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.* **23,** 837–848 (2009).

METHODS

Cell culture and generation of *POU5F1–GFP* reporter cell-line. The hESC lines H1 (WA-01, passage 28), H9 (WA-09, passage 26), HES2 (ES-02, passage 79), HES3 (ES-03, passage 97) and H1 *POU5F1*–GFP reporter cells (passage 56) were used for this study[1]. They were cultured feeder-free on Matrigel (BD)[1]. Condition medium used for culturing hESCs contained 20% KO serum replacement, 1 mM L-glutamine, 1% non-essential amino acids and 0.1 mM 2-mercaptoethanol and an additional 8 ng ml^{-1} of basic fibroblast growth factor (Invitrogen) supplemented to the hESC unconditioned medium. Medium was changed daily. The hESCs were subcultured with 1 mg ml^{-1} collagenase IV (Gibco) every 5–7 days.

A 3,064-base pairs upstream region of human *POU5F1* gene was cloned upstream of a *GFP* reporter gene, in replacement of the cytomegalovirus promoter of the N-EGFP plasmid with Geneticin (Gibco) drug selection marker. The *POU5F1–GFP* construct (2 μg) was transfected into H1 hESCs using 6 μl of FuGENE (Roche). Drug resistant colonies appeared after 2 weeks of drug selection.

Transfection, staining and imaging in 384-well plates. Plates (384-well, Grenier) were coated with 10 μl of matrigel for 30 min at 37 °C before removing the excess Matrigel. Pooled or individual siRNAs (5 μl of 500 nM; siGenome, Dharmacon) were printed on the plates and frozen at −20 °C before use. During reverse transfection, a master mix of 0.05 μl of Dharmafect1 (Dharmacon) transfection reagent and 4.95 μl of OptiMEM (Invitrogen) mix was added to each well of the siRNA plates and incubated for 20 min. Subsequently, 3,000 cells in 40 μl of conditioned medium with 10 μM Rock inhibitor (Calbiochem) were seeded in each well. Reagents and cells were dispensed onto the plate using a multidrop (Thermo Scientific).

Using the H1 *POU5F1–GFP* reporter cell-line, we carried out a pilot experiment to determine the optimized conditions (amount of transfection reagent and cell density) for reverse transfection. To enable high transfection efficiency, the cells were dissociated into single cells and reverse-transfected with the siRNA/transfection lipid complexes. A Z′ factor of more than 0.5 was obtained from this pilot screen, indicating a robust dynamic range between the positive (*GFP* siRNA) and negative (non-targeting siRNA) controls for a high throughput screen.

For the genome-wide screen, the whole genome Dharmacon SMARTpooled siRNA library targeting 21,121 human genes was printed on 67 Matrigel-coated 384-well plates where each well contained a mixture of four siRNAs targeting a single gene. On each plate, we included the negative controls (non-targeting siRNA) and positive controls (*GFP* siRNA and *POU5F1* siRNA) in the designated wells. After 4 days of transfection, the medium was replaced with 30 μl of 4% paraformaldehyde (Sigma). The cells were fixed for 15 min before washing with PBS. Hoechst 3342 (1:10,000, Invitrogen) in 0.1%Triton-X/1% BSA was added to each well and stained for 30 min. The cells were then washed once with PBS and covered in 30 μl PBS.

Cells were imaged with an IXU ultra confocal microscope (Research Instruments) at ×20 magnification and four pictures were taken per well. Integrated fluorescent intensity and nuclei number were quantified using MetaXpress Image Acquisition and Analysis software V1.7. Z′ factor was calculated for the primary screen based on the formula $Z' = 1 - 3(\sigma p + \sigma n)/(\mu p - \mu n)$ where σp is the standard deviation of the positive control, σn is the standard deviation of the negative control, μp is the mean of the positive control and μn is the mean of the negative controls. z-score was calculated using the formula $z = (X - \mu)/$s.d. where μ is the mean of the negative controls and s.d. is the standard deviation of the whole population. X is the sample value calculated based on the integrated fluorescent intensity/ number of cells. The Z′ factor for the entire screen was 0.76.

Retroviral production and human iPSC induction. pMXs retroviral plasmids that carry cDNAs of human *OCT4*, *SOX2*, *KLF4* and *c-MYC* genes were obtained from Addgene (plasmids 17217, 17218, 17219 and 17220, S. Yamanaka)[16]. cDNAs of human *PRDM14*, *NFRKB* and *YAP1* were cloned into pMX vector for retrovirus-mediated overexpression. Retroviruses were packaged using Pantropic Retroviral Expression System (Clontech) and concentrated with centrifugal filter devices (Millipore). MRC-5 cells obtained from ATCC were cultured in 15% FBS/DMEM. Confluent MRC-5 cells were split into 24 wells 1 day ahead and then transduced with equal amount of the retroviruses stock in presence of 4 μg ml^{-1} polybrene (Sigma). After 24 h, the cells were changed to fresh 15% FBS/DMEM medium, and then split from 1 well of a 24-well plate into 2 wells of a 6-well plate with pre-seeded CF-1 feeders the next day. The cultures were then maintained in human ESC culture medium and fed every 2 days. To expand and characterize hiPSCs, each emerged hESC-like colony was mechanically dissociated to small clumps and transferred into one 6-well plate with CF-1 feeder.

Reporter assays. A minimal *POU5F1* proximal promoter region (350 bp) was cloned into the pGL3 basic vector (Promega), driving the *luciferase* gene via the cloning site BglII/NcoI. The CR2 and CR4 fragments (550 and 500 bp, respectively) were cloned into the PGL3-*POU5F1* proximal promoter vector downstream of the *luciferase* gene via the cloning site BamHI/SalI. For the cloning of reporter vector used to test the functional domains of PRDM14, three copies of 30-bp CR2 consensus motif were synthesized and cloned into the XhoI/BglII site in front of the minimal promoter of pGL4.23 vector (Promega) in tandem. H1, HES2 and HES3 hESCs were transfected with the reporter constructs using FuGENE (Roche) and E14 mESCs and 293T cells using Lipofectamine 2000 (Invitrogen). Cells were harvested 48–60 h after transfection and the luciferase activities were quantified using the Dual-luciferase Reporter Assay System (Promega).

LETTER

Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast

Norman Pavelka[1]*, Giulia Rancati[1]*, Jin Zhu[1]*, William D. Bradford[1], Anita Saraf[1], Laurence Florens[1], Brian W. Sanderson[1], Gaye L. Hattem[1] & Rong Li[1,2]

Aneuploidy, referring here to genome contents characterized by abnormal numbers of chromosomes, has been associated with developmental defects, cancer and adaptive evolution in experimental organisms[1–9]. However, it remains unresolved how aneuploidy impacts gene expression and whether aneuploidy could directly bring about phenotypic variation and improved fitness over that of euploid counterparts. Here we show, using quantitative mass spectrometry-based proteomics and phenotypic profiling, that levels of protein expression in aneuploid yeast strains largely scale with chromosome copy numbers, following the same trend as that observed for the transcriptome, and that aneuploidy confers diverse phenotypes. We designed a novel scheme to generate, through random meiotic segregation, 38 stable and fully isogenic aneuploid yeast strains with distinct karyotypes and genome contents between 1N and 3N without involving any genetic selection. Through quantitative growth assays under various conditions or in the presence of a panel of chemotherapeutic or antifungal drugs, we found that some aneuploid strains grew significantly better than euploid control strains under conditions suboptimal for the latter. These results provide strong evidence that aneuploidy directly affects gene expression at both the transcriptome and proteome levels and can generate significant phenotypic variation that could bring about fitness gains under diverse conditions. Our findings suggest that the fitness ranking between euploid and aneuploid cells is dependent on context and karyotype, providing the basis for the notion that aneuploidy can directly underlie phenotypic evolution and cellular adaptation.

Whole-chromosome or segmental aneuploidy has been observed in a wide range of organisms and conditions, from pathogenic and experimental fungal species adapting to growth inhibition, to human diseases such as cancer and Down's syndrome[2–10], but it is still not known how aneuploidy affects gene expression and cellular physiology[1,11–13]. The budding yeast *Saccharomyces cerevisiae*, with its 16-chromosome complement, has been a useful experimental model for addressing this question at a fundamental level. Experimental evolution in yeast suggested a correlation between the emergence of aneuploidy and adaptive phenotypes in response to various perturbations[3–7,10]. Transcriptome profiling demonstrated that aneuploidy causes changes in messenger RNA levels that mostly scale with chromosome copy numbers, and well beyond for some genes[7,10,11]. In contrast, two recent studies concluded that aneuploidy decreases cellular fitness irrespective of the specific karyotype[11,12] and suggested that dosage compensation for proteins encoded on aneuploid chromosomes correlates with a common stress response, dubbed 'proteotoxic stress'[11]. This raises the following conundrum: if net protein expression levels are insensitive to chromosome stoichiometry, and aneuploidy inevitably impairs fitness, how might aneuploidy provide phenotypic variation and possibly fitness advantages under selective conditions?[13] We sought to resolve this conundrum by using a panel of aneuploid yeast strains with a wide range of karyotypes grown under diverse conditions. In particular, our experiments were designed to answer the following questions. First, can aneuploidy directly confer phenotypic variation and possibly improved fitness? And second, is the proteome proportionally affected by chromosome copy-number variation due to aneuploidy?

To generate fully isogenic and stable aneuploid strains, containing a wide range of chromosome stoichiometries and without the need for any genetic selection, we induced meiosis in yeast strains with an odd ploidy (3N or 5N), which produces aneuploid progenies at high frequencies[14,15] (Fig. 1a). To minimize other genetic variation, we generated the starting triploid and pentaploid progenitor strains by cycles of mating type switching and mating, from a single haploid S288c strain (Supplementary Fig. 1a). Absence of segmental chromosome abnormalities in the resulting polyploid progenitors was verified by array-based comparative genomic hybridization (aCGH), although the pentaploid strain had quickly lost one copy of chromosomes III and V (Supplementary Fig. 1b). To isolate aneuploid strains with stable karyotypes, we used a multistep approach (detailed in Supplementary Methods). In brief, each spore was first spread on a yeast extract peptone dextrose plate to form single colonies (Fig. 1a). From each resulting plate with uniform colony sizes 11 colonies were picked randomly and analysed by fluorescence-activated cell sorting (FACS) to identify those original spores producing colonies with uniform ploidy (Supplementary Fig. 2a, b). Karyotype stability was further verified after freezing and revival (Supplementary Fig. 2c). The karyotypes of the final aneuploid strains were determined by a new quantitative polymerase chain reaction (qPCR)-based assay that allows accurate karyotyping in high-throughput formats (Fig. 1b–f, Supplementary Methods and Supplementary Fig. 2d). A subset of the karyotypes was also confirmed by aCGH (Fig. 1b–f and Supplementary Fig. 3). To minimize the accumulation of single-nucleotide mutations, all aneuploid strains were passaged no more than three times between initial derivation and experimental use. Indeed, whole-genome resequencing of the five aneuploid strains used for the transcriptome and proteome analyses (see below) revealed an absence of mutations in coding regions that were not already present in the parental strains (Supplementary Information). The above procedure yielded 38 isogenic stable aneuploid strains (12.5% of spores analysed) with a ploidy between 1N and 3N, harbouring 35 distinct karyotypes, mostly with multiple chromosomes in aneuploidy (Supplementary Fig. 4). Except for chromosomes II and IX, all 16 yeast chromosomes were equally represented as aneuploid chromosomes in this collection (Fig. 1g). The chromosome number distributions were skewed towards the left (lower number) from those expected from random meiosis I segregation (Fig. 1h), suggesting that strains with larger numbers of aneuploid chromosomes are either less viable or karyotypically unstable.

Growth comparison of the aneuploid strains with isogenic 1N, 2N and 3N euploid strains in rich medium at 23 °C confirmed the previous observation that most aneuploid strains grew poorly in comparison
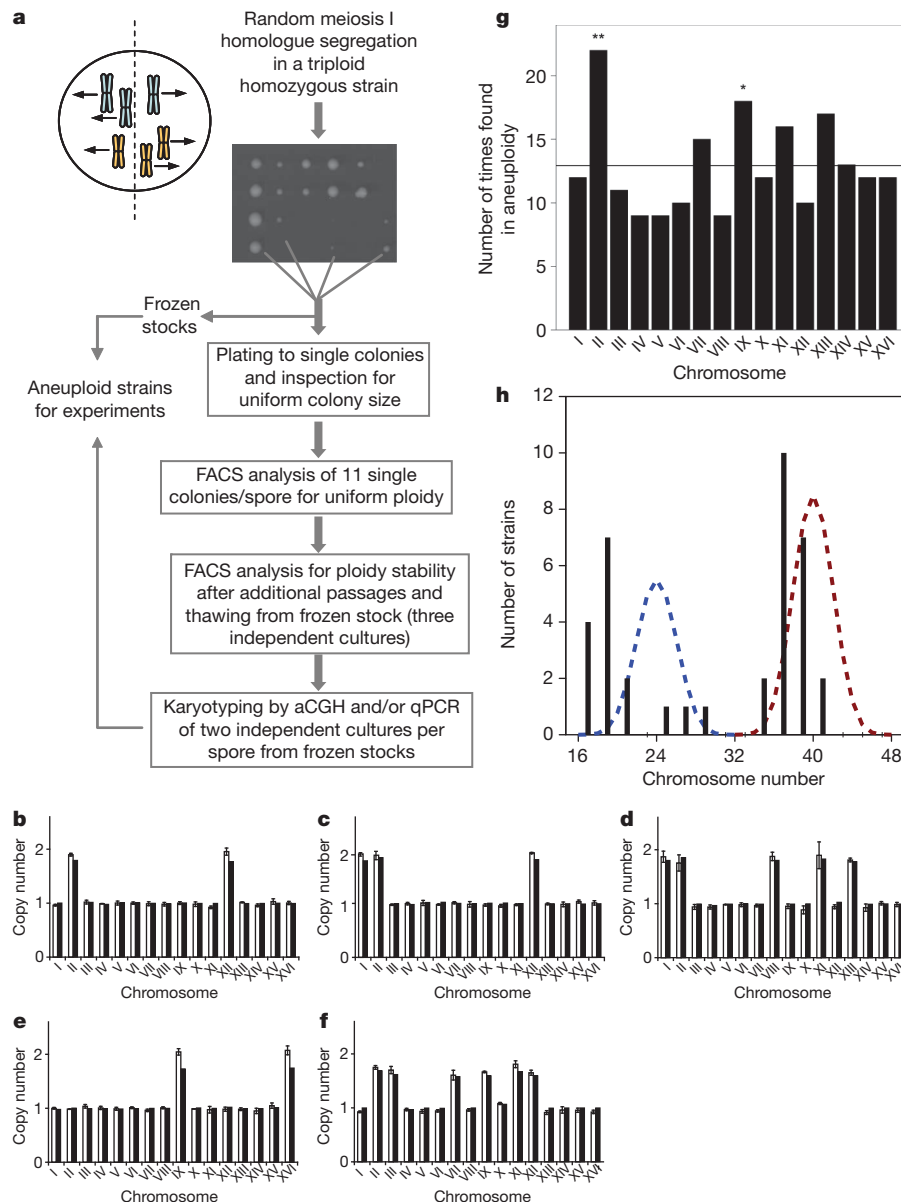
Figure 1 | Generation of aneuploid yeast strains. a, Sporulation of a homozygous triploid strain followed by karyotype stability tests of the meiotic progenies. **b–f**, Karyotypes of the five aneuploid strains (**b**, A2 (RLY4927); **c**, A3 (RLY4928); **d**, A13 (RLY4938); **e**, A14 (RLY4939); **f**, A15 (RLY4940)) used in Fig. 3, determined by qPCR (open bars; means ± s.d.) and aCGH (filled bars). **g**, Distribution of aneuploid chromosomes. The horizontal line represents the expectation number assuming uniform representation. Asterisk, $P < 0.05$; two

asterisks, $P < 0.001$ (P values calculated from a binomial distribution). **h**, Karyotype (total chromosome number) distribution across the aneuploid strain collection. Black lines, observed distribution (binned every two chromosomes); blue and red dashed lines, expected binomial distributions from random homologue segregation during triploid (blue line) and pentaploid (red line) sporulation.

with euploid controls[11], although a few aneuploid strains grew similarly to the euploids (Fig. 2a–c). Next we compared the growth under conditions divergent from that optimal for euploid yeast cells, including environmental perturbations, such as extreme temperature or pH or nutrient shortage, and the presence of chemotherapeutic or antifungal drugs (Supplementary Table 2). Strikingly, although under every condition most aneuploid strains grew more slowly than or as poorly as the euploids, under several conditions, especially those severely retarding euploid growth, some aneuploid strains showed improved growth in comparison with euploids (Fig. 2e–h and Supplementary Fig. 5). Karyotyping at the end of the growth assays confirmed persistent karyotype stability in most cases (Supplementary Information). For example, several aneuploid strains grew significantly better than euploid strains in rich medium at 16 °C (Fig. 2e, f) or in the presence of drugs such as rapamycin (an immunosuppressant and proposed anticancer agent),

bleomycin (a chemotherapeutic compound), thiolutin (an antibiotic) or fluconazole (an antifungal drug) at concentrations inhibitory to euploid growth (Fig. 2g, h, Supplementary Fig. 5 and Supplementary Table 3). The observed phenotypic diversity was unlikely to have been due to differences in mating type, because euploid strains with different mating types grew similarly under the tested conditions (Supplementary Fig. 6).

To understand how phenotypes relate to karyotypes, we clustered the aneuploid strains on the basis of karyotypic similarity (Fig. 2i) and clustered the conditions used in the phenotypic profiling on the basis of their effects on growth (Fig. 2j and Supplementary Methods). This analysis revealed that, first, several pairs of aneuploid strains with identical karyotypes (for example A26 and A27; A31 and A33) or similar karyotypes (for example A8 and A17; A1 and A16; A20 and A21) showed similar growth patterns across the different conditions.
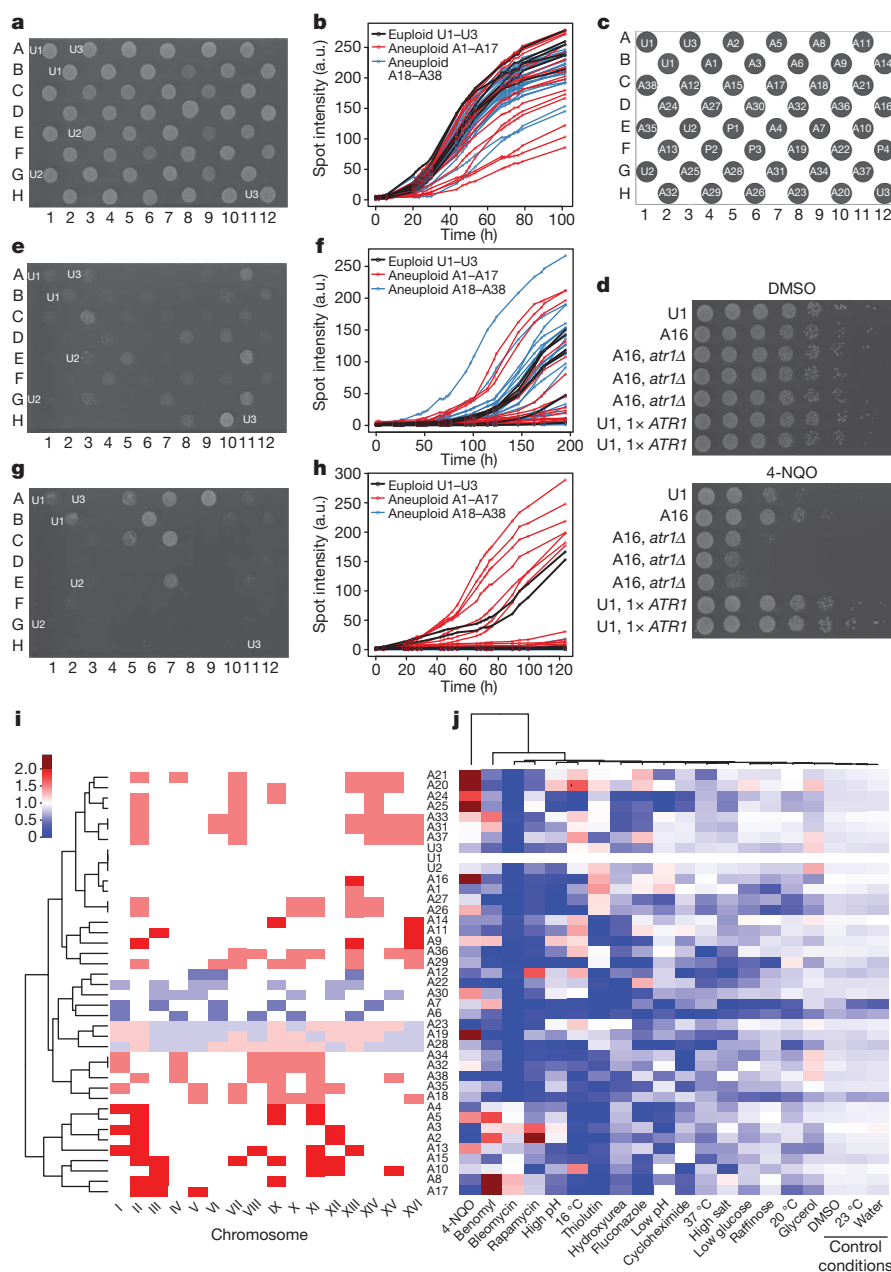
Figure 2 | Phenotypic profiling of aneuploid strains. a, b, e–h, Representative images (**a, e, g**) and growth curves (**b, f, h**) on YEPD plates at 23 °C (**a, b**), on YEPD plates at 16 °C (**e, f**) and in the presence of bleomycin (**g, h**). U1–U3 indicate haploid, diploid and triploid euploid control strains, respectively. a.u., arbitrary units. **c**, Strain positions. A1–A38 indicate aneuploid strains (see Supplementary Fig. 4 for their karyotypes); P1–P4 indicate four petite strains not studied further. **d**, One-copy-number increase in *ATR1* is required and sufficient to confer resistance to 0.4 μg ml⁻¹ 4-NQO. DMSO, dimethylsulphoxide. **i**, Clustering of strains based on karyotypic similarity. White, euploid chromosome number; red, gain over euploid number; blue, chromosome loss. **j**, Clustering of conditions used in phenotypic profiling based on the fitness relative to U1. White, growth similar to U1; red, fitness gain over U1; blue, fitness loss. The strains were ordered as in **i**. The scale bar applies to both **i** and **j**. Analysis details are given in Supplementary Methods.

Second, along each growth condition (column in Fig. 2j), divergent karyotypes can often be observed that showed improved growth compared with the haploid control. To pinpoint a specific mechanism linking a specific aneuploid karyotype to a specific fitness improvement, we noticed that strain A16, resistant to 4-nitroquinoline-*N*-oxide (4-NQO, a tumorigenic compound), had only chromosome XIII in aneuploidy. Chromosome XIII harbours the gene *ATR1*, encoding a transporter protein known to confer 4-NQO hyper-resistance when overexpressed[16]. We confirmed that an extra copy of chromosome XIII in strain A16 led to a proportional increase in mRNA expression of *ATR1* in both the presence and absence of 4-NQO (Supplementary Fig. 7a). Deletion of the extra copy of *ATR1* from the 4-NQO-resistant aneuploid strains restored the expression of *ATR1* to levels comparable

to those of the euploid (Supplementary Fig. 7a) and abolished their resistance to 4-NQO (Fig. 2d and Supplementary Fig. 7b, c), demonstrating that the increased *ATR1* copy number was required for 4-NQO resistance. Furthermore, introducing one extra copy of the *ATR1* gene expressed under its own promoter into haploid or diploid euploid strains was sufficient to confer resistance to 4-NQO (Fig. 2d and Supplementary Fig. 7b, c).

Immunoblot analysis of disomic yeast strains suggested that most proteins encoded on aneuploid chromosomes may be dosage compensated[11], raising a question as to how aneuploidy might confer large phenotypic variation, as observed above. To investigate this further, we performed parallel RNA microarray and quantitative proteomics analysis on five aneuploid strains of the same mating type but with

different growth rates and chromosome stoichiometries. MudPIT (multidimensional protein identification technology) analysis[17] on the soluble fraction of whole-cell extracts identified about 2,000 different gene products per strain, representing about 33% of the yeast proteome, with highly reproducible quantification of protein abundances across the biological replicates and broad coverage of all cellular components (Supplementary Fig. 8 and Supplementary Table 4). This analysis revealed that strains with similar karyotypes tend to have similar global proteomic changes (Supplementary Fig. 8a) and that changes in chromosome copy numbers resulting from aneuploidy lead to proportional changes in the chromosomal average protein expression, following the same trend as the transcriptome (Fig. 3a–c and Supplementary Figs 9 and 10a), indicating a direct gene dosage effect on the proteome. Dosage compensation was also minimal and in most cases insignificant for core complex proteins[18] encoded on aneuploid chromosomes (Supplementary Fig. 10b).

The mRNA and protein levels of individual genes were modestly, although significantly, correlated even after correcting for chromosomal copy number effect (Fig. 3d and Supplementary Fig. 11), which is consistent with a recent proteomic analysis of haploid versus diploid yeast[19]. However, among the genes with expression changes two standard

deviations away from their chromosomal average (referred to as 'outlier' genes), only a fraction (about 3–14%) were common between the microarray and the proteomics data sets (Fig. 3d and Supplementary Fig. 11). As a result, mRNA outliers and protein outliers were enriched for distinct classes of biological processes, and no specific class of genes was consistently found to be significantly enriched across all aneuploid strains (Fig. 3e, f and Supplementary Fig. 12a, b), suggesting a lack of a common gene expression response to aneuploidy. Because genes of the 'response to stress' category were not enriched in either the transcriptome or the proteome in any of our aneuploid strains, in contrast with the conclusion of the previous study[11], we further performed a more stringent analysis by considering only outlier genes expressed more than three standard deviations from the chromosomal average in our microarray data. Enrichment for either 'response to stress' (Gene Ontology) or 'environmental stress response' genes[20] was found in three of the five aneuploid strains, but this enrichment was not correlated with either growth rates or the number of aneuploid chromosomes (Supplementary Fig. 12c). Taken together, the above results indicate that aneuploidy has global and complex effects at both the transcriptome and proteome levels, and that an increase in stress gene expression is not a obligate property of aneuploid strains.
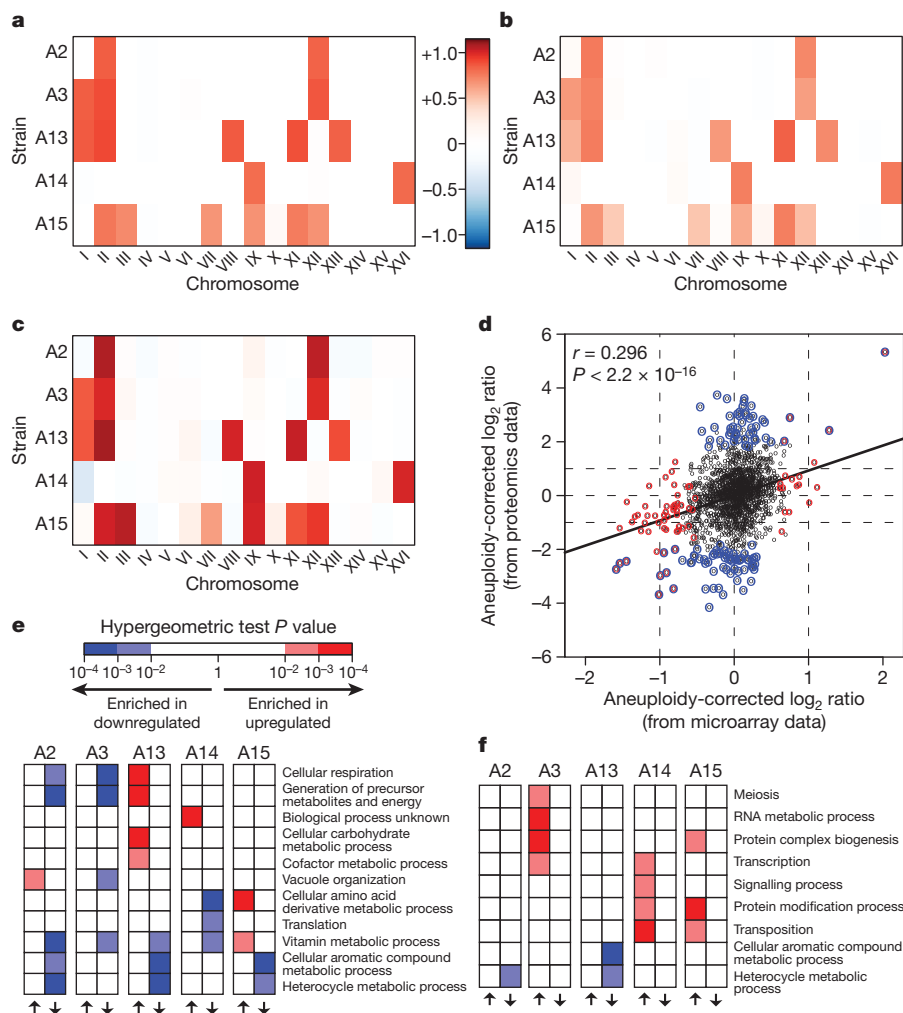
Figure 3 | Effects of aneuploidy on the proteome. a–c, Heat maps of chromosome stoichiometry (**a**, aCGH data; normalized DNA log$_2$ ratio (aneuploid/euploid); Fig. 1b–f), average mRNA level (**b**, microarray data; normalized mRNA log$_2$ ratio (aneuploid/euploid)) and average protein level (**c**, proteomics data; normalized protein log$_2$ ratio (aneuploid/euploid); see Supplementary Fig. 9) per chromosome of the five aneuploid strains compared with U1. **d,** Correlation between protein expression and gene expression changes in strain A2 relative to haploid euploid strain U1 (for other strains, see Supplementary Fig. 11). Outlier mRNAs and proteins (defined as in Supplementary Information) are highlighted in red and blue, respectively. **e, f,** Subset of GO-Slim (a cut-down version of gene ontologies containing a subset of the GO terms) analysis applied to outlier genes from microarray (**e**) and proteomics (**f**) data sets (see Supplementary Methods for details). Complete results are given in Supplementary Fig. 12a, b. P values were calculated from hypergeometric tests.

Our analysis of a large set of isogenic and stable aneuploid yeast strains with broad chromosome stoichiometries demonstrates that aneuploidy directly confers phenotypic variation and sometimes growth advantage under conditions that are suboptimal for euploid cells. These observations suggest that aneuploidy does not inevitably result in growth impairment, but rather that the impact of aneuploidy on cellular fitness depends on both karyotype and conditions. The difference between our findings and the previous observation of a common stress signature and proliferative disadvantage across disomic yeast strains[11] may be that in our study the naturally stable, multiple-chromosome aneuploidy resulted in fewer protein expression imbalances than single-chromosome aneuploidy maintained through continuous selection with drugs and nutrient markers. Furthermore, our proteomic analysis, performed in quadruplicate and quantifying thousands of proteins encoded on aneuploid chromosomes, in contrast with just 16 proteins analysed by immunoblotting in the previous work, revealed a wholesale chromosome dosage effect on the proteome, consistent with a recent report of effects of gene copy number on protein levels[21]. These findings suggest that aneuploidy is a large-effect mutation profoundly altering gene expression at the functional level. Under conditions that euploid cells are well adapted to, the large phenotypic effects caused by aneuploidy are likely to cause a decrease in fitness that could lead to the rapid clearance of most aneuploid cells from the population. However, under a strong selective pressure due to adverse environmental changes or clinical drug treatments, the rise of aneuploidy, readily achieved through erroneous mitosis, can be a highly effective mechanism to generate phenotypic variation and rapid adaptation.

METHODS SUMMARY

Generation of a collection of isogenic aneuploid yeast strains. Homozygous triploid and pentaploid strains were generated as described in Supplementary Fig. 1a. Aneuploid strains were generated by sporulation of the above polyploid strains, followed by karyotype stability tests and determination as described in Fig. 1a and Supplementary Fig. 2. Strains and plasmid are listed in Supplementary Information.

Karyotyping. aCGH, performed as described previously[7], and qPCR were used for karyotyping. qPCR assays were designed with primers in non-coding regions on each chromosome arm (primer sequences are listed in Supplementary Table 1). DNA samples were prepared by alkaline lysis, and qPCR reactions were performed in 384-well plates with a BioMek FX (Beckman Coulter) to assemble 10-µl reactions and an ABI 7900HT (Applied Biosystems) for cycling. Chromosome copy numbers were determined with a modified $\Delta\Delta C_t$ method (Supplementary Methods).

Phenotypic profiling. Equal amounts (D_{600}) of aneuploid and euploid control cultures were spotted, with the Biomek FX robot, onto OmniTrays containing various solid media and grown under conditions listed in Supplementary Table 2. OmniTrays representing three biological replicates of each tested condition were scanned on a Hewlett-Packard ScanJet 4070 desktop scanner. Growth data were obtained by automated spot detection and intensity measurements.

Quantitative whole-genome proteomics. Whole-cell lysates were prepared from 50 ml of cycling yeast cultures by bead-beating. High-speed supernatants were collected and precipitated. Chromatography and mass spectrometry analysis were performed as described previously[22]. The tandem mass spectrometry data sets were searched using SEQUEST[23] against a database of 11,986 sequences, consisting of 5,816 S. cerevisiae non-redundant proteins (NCBI), 177 contaminants and 5,993 decoy sequences. Relative protein levels were determined by calculating distributed normalized spectral abundance factors[24].

Statistical analysis. All statistical analyses were performed in the R environment[25] using standard packages and custom scripts.

1. Torres, E. M., Williams, B. R. & Amon, A. Aneuploidy: cells losing their balance. *Genetics* **179,** 737–746 (2008).
2. Weaver, B. A. & Cleveland, D. W. Does aneuploidy cause cancer? *Curr. Opin. Cell Biol.* **18,** 658–667 (2006).
3. Selmecki, A., Forche, A. & Berman, J. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans. Science* **313,** 367–370 (2006).
4. Polakova, S. et al. Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata. Proc. Natl Acad. Sci. USA* **106,** 2688–2693 (2009).
5. Dunham, M. J. et al. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA* **99,** 16144–16149 (2002).
6. Gresham, D. et al. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* **4,** e1000303 (2008).
7. Rancati, G. et al. Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135,** 879–893 (2008).
8. Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A. & Berman, J. An isochromosome confers drug resistance *in vivo* by amplification of two genes, *ERG11* and *TAC1. Mol. Microbiol.* **68,** 624–641 (2008).
9. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet.* **5,** e1000705 (2009).
10. Hughes, T. R. et al. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genet.* **25,** 333–337 (2000).
11. Torres, E. M. et al. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317,** 916–924 (2007).
12. Williams, B. R. et al. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322,** 703–709 (2008).
13. Pavelka, N., Rancati, G. & Li, R. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. *Curr. Opin. Cell Biol.* advance online publication doi:10.1016/j.ceb.2010.06.003 (23 July 2010).
14. Parry, E. M. & Cox, B. S. The tolerance of aneuploidy in yeast. *Genet. Res.* **16,** 333–340 (1970).
15. St Charles, J., Hamilton, M. L. & Petes, T. D. Meiotic chromosome segregation in triploid strains of *Saccharomyces cerevisiae. Genetics* doi:10.1534/genetics.110.121533 (10 August 2010).
16. Mack, M. et al. Genetic characterization of hyperresistance to formaldehyde and 4-nitroquinoline-N-oxide in the yeast *Saccharomyces cerevisiae. Mol. Gen. Genet.* **211,** 260–265 (1988).
17. Washburn, M. P., Wolters, D. & Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19,** 242–247 (2001).
18. Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).
19. de Godoy, L. M. et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455,** 1251–1254 (2008).
20. Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11,** 4241–4257 (2000).
21. Springer, M., Weissman, J. S. & Kirschner, M. W. A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* **6,** 368 (2010).
22. Florens, L. & Washburn, M. P. Proteomic analysis by multidimensional protein identification technology. *Methods Mol. Biol.* **328,** 159–175 (2006).
23. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).
24. Zhang, Y., Wen, Z., Washburn, M. P. & Florens, L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82,** 2272–2281 (2010).
25. Ihaka, R. & Gentleman, R. R. A language for data analysis and graphics. *J. Comput. Graph. Statist.* **5,** 299–314 (1996).

LETTER

The mechanism of retroviral integration from X-ray structures of its key intermediates

Goedele N. Maertens[1], Stephen Hare[1] & Peter Cherepanov[1]

To establish productive infection, a retrovirus must insert a DNA replica of its genome into host cell chromosomal DNA[1,2]. This process is operated by the intasome, a nucleoprotein complex composed of an integrase tetramer (IN) assembled on the viral DNA ends[3,4]. The intasome engages chromosomal DNA within a target capture complex to carry out strand transfer, irreversibly joining the viral and cellular DNA molecules. Although several intasome/transpososome structures from the DDE(D) recombinase superfamily have been reported[4–6], the mechanics of target DNA capture and strand transfer by these enzymes remained unclear. Here we report crystal structures of the intasome from prototype foamy virus in complex with target DNA, elucidating the pre-integration target DNA capture and post-catalytic strand transfer intermediates of the retroviral integration process. The cleft between IN dimers within the intasome accommodates chromosomal DNA in a severely bent conformation, allowing widely spaced IN active sites to access the scissile phosphodiester bonds. Our results resolve the structural basis for retroviral DNA integration and provide a framework for the design of INs with altered target sequences.

To learn how the retroviral integration machinery engages chromosomal DNA, we co-crystallized the prototype foamy virus (PFV) intasome with a model target DNA (tDNA) construct (Fig. 1a), which was designed on the basis of the PFV integration site consensus[7,8]. Addition of Mg^{2+} allowed strand transfer to occur during crystallization experiments (Supplementary Fig. 1), resulting in crystals of the post-catalytic strand transfer complex (STC), while pre-catalytic target capture complex (TCC) crystals were obtained in the absence of the essential catalytic metal (TCC_{Apo}). Furthermore, using a viral DNA mimic lacking the reactive 3′-hydroxyl group enabled us to grow crystals of the catalytically trapped complex (TCC_{ddA}) in the presence of Mg^{2+}, which considerably extended their diffraction limit. The STC, TCC_{Apo}, and TCC_{ddA} structures were refined to 2.81 Å, 3.32 Å and 2.97 Å resolution, respectively (Supplementary Table 1, Supplementary Fig. 2).

As predicted[4], the tDNA is accommodated within the cleft between the halves of the symmetric intasome (Fig. 1b, c and Supplementary Movie 1). The intasome does not rearrange (Supplementary Fig. 3) but induces severe bending of the tDNA duplex (Fig. 2a and Supplementary Movie 2). Crucially, this binding mode provides the active sites of the inner IN subunits, separated as far as 26.5 Å in the TCC structures, with direct access to the scissile phosphodiester bonds within tDNA (Supplementary Figs 2a–c, 4 and 5). Superposition of the TCC_{ddA} structure and the Mn^{2+}-bound form of the intasome[4] positions the 3′-hydroxyl group of viral DNA, coordinated to metal B of the active site, for in-line S_N2 nucleophilic substitution at the phosphorus atom of the scissile phosphodiester in tDNA (Fig. 2b). Notably, the sugar moiety of the tDNA nucleotide at the site of strand transfer (cytosine 0) adopts different conformations in the TCC and STC structures. Primarily due to a ~110° rotation around the deoxyribose C4–C5 bond, the viral DNA-tDNA phosphodiester shifts from its pre-strand transfer position by 2.3 Å and is thereby ejected from the active site in the STC structure (Fig. 2b). This conformational change is expected to prevent a reversal of the strand transfer reaction.

Overall, the conformations of the synapsed DNA molecules within the TCC and STC structures are similar (Fig. 2a and Supplementary Movie 2), and will be discussed in the context of the STC. At the centre of the integration site, the major groove of the target is widened to 26.3 Å, and the minor groove is compressed to 9.6 Å with a 55° negative roll between base pairs involving thymine 1 and adenine 2, resulting in
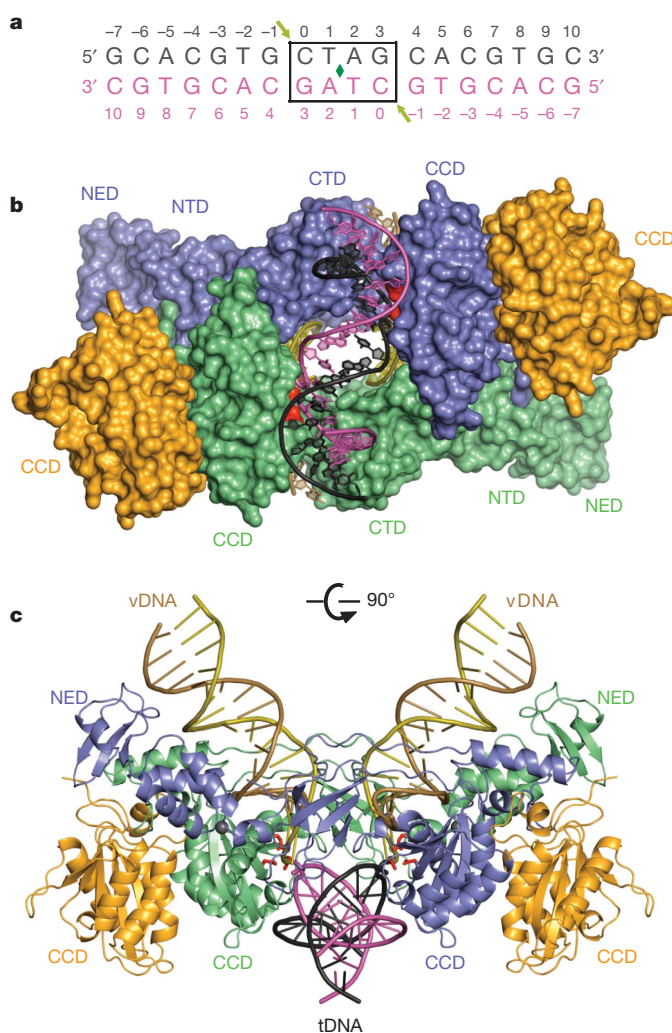
Figure 1 | Crystal structure of the PFV STC. a, Sequence of the model tDNA. The arrows point to the sites of strand transfer with a 4-base pair (bp) stagger (boxed). Nucleotide numbers are indicated for both strands. The green diamond indicates the 2-fold symmetry axis. **b**, **c**, Surface (**b**) and cartoon (**c**) representations of the STC crystal structure, viewed along or perpendicular to the crystallographic two-fold axis, respectively. IN active site carboxylates are shown in red. NTD, CCD, CTD and NTD extension domain (NED) of IN as well as viral (v) and target (t) DNAs are indicated.

[1]Division of Infectious Diseases, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK.
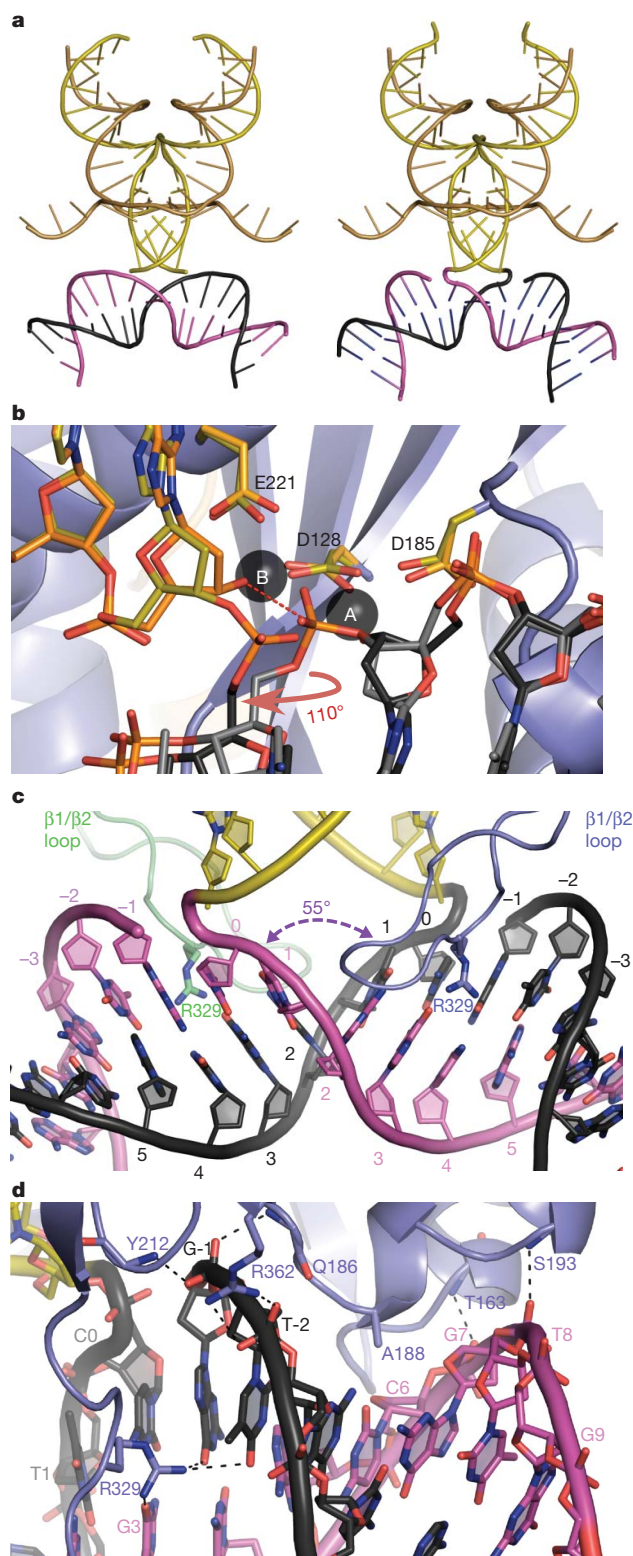
Figure 2 | Details of DNA conformations, recognition, and active site mechanics during strand transfer. a, Cartoon representations of the viral and target DNAs in the TCC$_{Apo}$ (left panel) and STC (right panel) structures. **b**, Mechanics of DNA strand transfer. Superposition of common elements within TCC$_{ddA}$, STC, and Mn^{2+}-loaded intasome[4] (PDB ID 3L2S). Carbon atoms of viral DNA and IN active site residues are orange (3L2S) and yellow (STC). **c**, Proximal view of the integration site in the STC structure. **d**, IN interactions with tDNA. Dashed lines indicate hydrogen bonds.

the unstacking of the two consecutive base pairs (Fig. 2c). Remarkably, this severe DNA kinking does not involve direct interactions between unstacked base pairs and protein. The TCC and STC are stabilized by eight rigid hydrogen bonds between amide groups of the protein main chain (residues Thr 163, Gln 186, Ser 193 and Tyr 212 from each inner IN monomer) and the tDNA phosphodiester backbone in addition to a pair of salt bridges involving inner chain carboxy-terminal domain (CTD) Arg 362 residues (Fig. 2d). As can be expected from the relatively low degree of sequence selectivity of retroviruses for chromosomal DNA[9–11], interactions between IN and tDNA bases are sparse. Nonetheless, two sites of close contact were identified within the TC/STC structures. Firstly, the side chain of Arg 329, based on the loop connecting IN CTD β1 and β2 strands (β1/β2 loop), is hydrogen-bonded to guanine 3, guanine −1, and thymine −2 bases within the expanded major groove of the tDNA (Fig. 2c, d and Supplementary Fig. 2d). Secondly, the methyl group of Ala 188, at the beginning of α2 helix of the inner IN chain catalytic core domain (CCD), is involved in a van der Waals interaction with the O2 atom of cytosine 6 within the minor groove (Fig. 2d).

Owing to their naturally low extent of base stacking, pyrimidine (Y) – purine (R) dinucleotide steps are known to be the most flexible, followed by YY (or RR) and, the least deformable, RY steps[12]. In concordance, PFV integration sites are enriched in YR dinucleotides at positions 1 and 2 (Supplementary Fig. 6). The length of the β1/β2 loop and the presence of Arg at an equivalent position to 329 are invariant among spumaviral INs, underscoring the importance of the observed interactions. In PFV IN, substitution of Arg 329 for Ser, a residue with a smaller side chain, is expected to abolish hydrogen bonding with the tDNA bases as well as reduce the geometric fit between the intasome and the functional tDNA conformation. While not affecting intasome assembly and strand transfer activity (Supplementary Fig. 7), the R329S mutation increased the bias of PFV IN against the rigid RY dinucleotides at positions 1 and 2 of the resulting integration sites (Fig. 3a, b and Supplementary Fig. 6), confirming the role of Arg 329 in tDNA bending. Another mutant, R329E greatly reduced strand transfer activity of the intasome (Supplementary Fig. 7). Sequencing of the residual R329E strand transfer products revealed a striking preference for guanosine at position 4 (and a symmetric preference for cytosine at position −1) of these integration sites ($P < 10^{-32}$) (Fig. 3c). Interaction of the mutant Glu residue with a cytidine base[13] at position −1 of the integration site probably accounts for the marked tDNA sequence preferences of the R329E intasome. Strong selectivity towards chromosomal DNA should limit possible integration sites and consequently reduce viral fitness. Arg, a residue with a flexible, protonated side chain, is able to form a plethora of hydrogen bond interactions within the major groove of DNA[13], helping to offset the energetic penalty associated with tDNA bending, while introducing only minor sequence preferences for positions 4 and 5 (Fig. 3a).

In retroviral INs, the positions equivalent to PFV 188 are invariably occupied by small amino acids, typically Ala, Pro or Ser, suggesting that a close contact between the CCD α2 helix and the minor groove of tDNA is a common feature of retroviral TC/STCs. In line with prior observations[14,15], substitution of PFV IN Ala 188 (a residue structurally equivalent to HIV-1 Ser 119) for Asp ablated strand transfer activity (Supplementary Fig. 7). A less drastic mutation, A188S, did not affect the level of strand transfer activity, but yielded a significant bias for adenosine at tDNA position 6 ($P = 10^{-4}$) (Fig. 3d), possibly owing to the hydrogen bonding of the mutant side chain to N3 of the adenine base.

One puzzling feature of the PFV integration site consensus is a pronounced bias against thymidine at position 0 and the symmetrical avoidance of adenosine at position 3 (Fig. 3a). Furthermore, the bias against integration immediately upstream of a thymidine appears to be a common feature of retroviruses[9–11]. Modelling a T•A base pair at the site of integration reveals that the C5-methyl group of thymine would
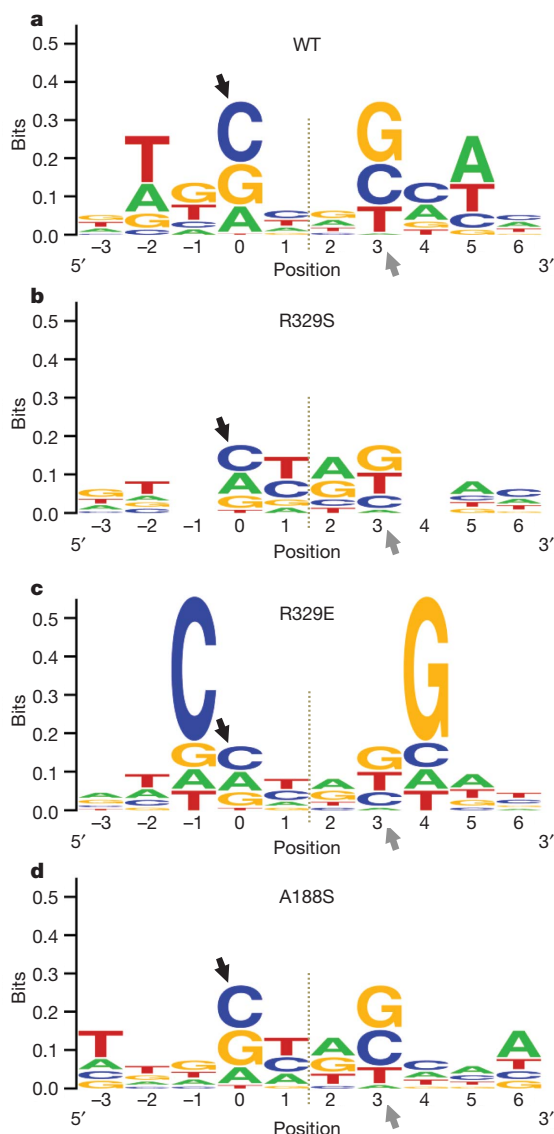
Figure 3 | Sequence analysis of strand transfer reaction products.
a–d, Sequence logos representing nucleotide frequencies at PFV integration sites *in vitro* using wild-type (**a**), R329S (**b**), R329E (**c**) or A188S (**d**) INs. Target DNA sequences as joined to the viral DNA ends were used in the alignments. Arrowheads indicate sites of strand transfer, and dotted lines symmetry axes of the integration sites; the strict two-fold symmetry is imposed by the use of intasomes containing pairs of U5 viral DNA ends. At each position, the height of a letter is proportional to the observed frequency of the corresponding nucleotide, and the height of each letter stack corresponds to the level of conservation, represented as information content (y-axis, bits).

hinder the target phosphodiester, approaching it at 3.4 Å in the STC structure (Supplementary Fig. 8) and thus probably interfering with the mechanics of transesterification.

The conformation of tDNA within the STCs is fully consistent with previous observations that pre-bending of tDNA promotes integration[16,17]. Of note, *in vitro* PFV IN appears more selective for flexible target sequences compared with what is observed during viral infections (Supplementary Fig. 6). We speculate that chromosomal DNA packaging into nucleosomes and the presence of other DNA bending factors in cells contribute to this subtle difference. In the context of integration into chromatin, the N-terminal domains (NTDs) and CTDs of the outer IN subunits, disordered in our crystals, might interact with nucleosomal DNA and/or the histone octamer. The TCC and STC structures we present here elucidate the mechanism for retroviral DNA integration and indirect recognition of the optimal

tDNA sequence. They furthermore provide a framework for the design of INs with altered target sequences and will boost ongoing efforts to create site-specific retroviral vector systems for applications in gene therapy[18,19]. The PFV intasome[4] as well as the TCC and STC structures detailed here will enable the building of reliable models for the respective intermediates of the HIV integration process[20] to aid improvement of the existing antiviral strategies and to hasten discovery of novel approaches to blocking viral replication.

METHODS SUMMARY

PFV intasomes, assembled with wild-type (WT), full-length IN and a mimic of the pre-processed U5 viral DNA containing or lacking the reactive 3′-hydroxyl group, were co-crystallized with a self-complementary tDNA oligonucleotide 5′-CCCGAGGCACGTGCTAGCACGTGCCTCGGG. Inclusion of the natural 3′-hydroxyl in the viral DNA construct allowed strand transfer to occur during crystallization in the presence of $MgCl_2$ (Supplementary Fig. 1). Although the intasome can engage the tDNA construct at multiple sites, the constraints imposed by crystal symmetry and lattice contacts probably accounted for selective crystallization of the symmetric complex (compare lanes 3 and 4 in Supplementary Fig. 1c). In both TCC and STC crystals, the asymmetric units contained half of the intasome structure (IN chains A and B, viral DNA strands C and D) and one strand of tDNA (chain T); the complete biological assemblies were generated by crystallographic two-fold symmetry operations. Sixteen and 18 base pairs of tDNA could be built into the TCC and STC electron density maps, respectively. Similar to the original PFV intasome crystals[4], the N- and C-terminal domains of the outer IN subunits (chains B and B′) were disordered and are not present in the final models. Crystallographic and refinement statistics are summarized in Supplementary Table 1. The final models had good geometry with 96.7% and 0% (STC), 92.9% and 0.4% (TCC_{ddA}), and 91.8% and 1.2% (TCC_{Apo}) of amino acid residues in most preferred and disallowed regions of the Ramachandran plot, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Craigie, R. in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M) 613–630 (ASM Press, 2002).
2. Lewinski, M. K. & Bushman, F. D. Retroviral DNA integration--mechanism and consequences. *Adv. Genet.* **55,** 147–181 (2005).
3. Li, M., Mizuuchi, M., Burke, T. R. Jr & Craigie, R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* **25,** 1295–1304 (2006).
4. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464,** 232–236 (2010).
5. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. & Rayment, I. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289,** 77–85 (2000).
6. Richardson, J. M., Colloms, S. D., Finnegan, D. J. & Walkinshaw, M. D. Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* **138,** 1096–1108 (2009).
7. Trobridge, G. D. *et al.* Foamy virus vector integration sites in normal human cells. *Proc. Natl Acad. Sci. USA* **103,** 1498–1503 (2006).
8. Valkov, E. *et al.* Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res.* **37,** 243–255 (2009).
9. Berry, C., Hannenhalli, S., Leipzig, J. & Bushman, F. D. Selection of target sites for mobile DNA integration in the human genome. *PLOS Comput. Biol.* **2,** e157 (2006).
10. Holman, A. G. & Coffin, J. M. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl Acad. Sci. USA* **102,** 6103–6107 (2005).
11. Wu, X., Li, Y., Crise, B., Burgess, S. M. & Munroe, D. J. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79,** 5211–5214 (2005).
12. Johnson, R. C., Stella, S. & Heiss, J. K. in *Protein-nucleic Acid Interactions* (eds Rice, P. A. & Correll, C. C.) Ch. 8, 176–220 (RSC Publishing, 2008).
13. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29,** 2860–2874 (2001).
14. Konsavage, W. M. Jr, Burkholder, S., Sudol, M., Harper, A. L. & Katzman, M. A substitution in Rous sarcoma virus integrase that separates its two biologically relevant enzymatic activities. *J. Virol.* **79,** 4691–4699 (2005).
15. Harper, A. L., Sudol, M. & Katzman, M. An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. *J. Virol.* **77,** 3838–3845 (2003).
16. Müller, H. P. & Varmus, H. E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13,** 4704–4714 (1994).
17. Pruss, D., Bushman, F. D. & Wolffe, A. P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl Acad. Sci. USA* **91,** 5913–5917 (1994).

18. Lim, K. I., Klimczak, R., Yu, J. H. & Schaffer, D. V. Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties. *Proc. Natl Acad. Sci. USA* **107,** 12475–12480 (2010).
19. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature Biotechnol.* **25,** 1298–1306 (2007).
20. Krishnan, L. *et al.* Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc. Natl Acad. Sci. USA* **107,** 15910–15915 (2010).

METHODS

Intasome preparation, crystallization, and structure determination. Full-length, wild-type PFV IN and its mutant forms were produced according to established procedures[4]. Synthetic DNA was purchased from Eurogentec and Midland Certified. Donor DNA was obtained by annealing HPLC-purified synthetic oligonucleotides 5′-TGCGAAATTCCATGACA (reactive strand) or its 3′-deoxy derivative 5′-TGCGAAATTCCATGAC[2′,3′-ddA] and 5′-ATTGTC ATGGAATTTCGCA (non-transferred strand). The intasomes, assembled by dialysis of IN/donor DNA mixtures as previously described[4] and were purified by chromatography through HiLoad 16/60 Superdex-200 column (GE Healthcare) in 320 mM NaCl, 20 mM bis-tris propane (BTP)-HCl at pH 7.45. Intasome preparations, mixed with 1.5-molar excess of self-annealed tDNA oligonucleotide 5′-CC CGAGGCACGTGCTAGCACGTGCCTCGGG, were dialysed overnight against excess of 225 mM NaCl, 2 mM dithiothreitol (DTT), 25 μM ZnCl₂, 20 mM BTP-HCl, pH 7.45 and concentrated to 6 mg ml⁻¹ using Vivaspin-4 devices with a 10-kDa cut-off.

Hanging drop vapour diffusion crystallization experiments were set up by mixing 1 μl of protein-DNA complex solution with 1 μl reservoir solution. The reservoir contained 20% (±)-2-methyl-2,4-pentanediol (MPD), 40 mM MgCl₂ and 50 mM sodium cacodylate, pH 5.4 (STC); 38% PEG400, 180 mM Li₂SO₄, 20 mM MgCl₂, 100 mM Tris-HCl, pH 7.5 (TCC_ddA); or 34% PEG400, 180 mM Li₂SO₄ and 100 mM Tris-HCl, pH 7.5 (TCC_Apo). Crystals, grown at 18 °C, appeared within 2–3 days and reached a size of 100–400 μm within 2–4 weeks. Crystals, cryoprotected in 30% MPD, 3 mM MgCl₂, 150 mM NaCl and 30 mM sodium cacodylate–NaOH, pH 5.9 (STC); 40% PEG400, 150 mM NaCl, 150 mM Li₂SO₄, 15 mM MgCl₂ and 100 mM Tris-HCl, pH 7.4 (TCC_ddA); or 35% PEG400, 150 mM NaCl, 150 mM Li₂SO₄ and 100 mM Tris-HCl, pH 7.5 (TCC_Apo), were frozen by quick submersion in liquid nitrogen.

Diffraction data, collected at beam line I02 of the Diamond Light Source at 100 K, were integrated with XDS[21], merged and scaled using SCALA from the CCP4 suite[22]. Crystals belonged to space group $P4_12_12$ and were nearly isomorphous to crystals obtained in the absence of tDNA[4]. The structures were solved by isomorphous replacement using intasome coordinates from PDB entry 3L2R (ref. 4), and the tDNA strands were built into the resulting $F_o - F_c$ difference maps. The final structures, refined using Refmac[23], were validated with MolProbity[24]. The models had good geometry with 96.7% and 0% (STC), 92.9% and 0.4% (TCC_ddA), and 91.8% and 1.2% (TCC_Apo) of amino acid residues in most preferred and disallowed regions of the Ramachandran plot, respectively. X-Ray data collection and refinement statistics are given in Supplementary Table 1; examples of final weighted $2F_o - F_c$ electron density maps are given in Supplementary Fig. 2. Structure figures were generated using PyMOL (http://www.pymol.org).

Strand transfer assays and integration product sequencing. Strand transfer assays used WT and mutant intasomes purified by size exclusion chromatography through a Superdex 200HR 10/30 column (GE Healthcare) in 320 mM NaCl, 20 mM

BTP-HCl, pH 7.45 (Supplementary Fig. 7) and quantified by spectrophotometry at 260 nm. Strand transfer assays contained 300 ng supercoiled pGEM9-Zf(−) (Promega) tDNA and 1.6 pmol or 0.4 pmol intasome in 40 μl of 115 mM NaCl, 5 mM MgCl₂, 1 mM dithiothreitol, 4 μM ZnCl₂, 25 mM BTP-HCl, pH 7.45. Reactions were incubated at 37 °C for 30 min and stopped by addition of 0.5% SDS and 25 mM EDTA. DNA products deproteinized by digestion with 20 μg proteinase K for 30 min at 37 °C and precipitated with ethanol were separated in 1.5% agarose gels and visualized by staining with ethidium bromide. For sequence analyses[8], concerted strand transfer products isolated from agarose gels were treated with phi29 DNA polymerase (New England Biolabs) in the presence of 450 μM each dNTP, 5′-phosphorylated using T4 polynucleotide kinase (New England Biolabs) and ligated with a blunt-ended DNA fragment spanning the Tn5 kanamycin resistance gene flanked by KpnI restriction sites. The kanamycin resistance cassette was generated by PCR using primers 5′-GGCGGGGTACCAGAAAGCAGGTAGCTT GCAGTGG and 5′-GGCGGGGTACCCGAAGAACTCCAGCATGAGATCC (KpnI sites underlined) and pCP15 (ref. 25) as a template. *Escherichia coli* TOP10 cells (Invitrogen), transformed with the ligation products, were selected with 35 μg ml⁻¹ kanamycin. Plasmids isolated from individual clones were analysed by digestion with KpnI (New England Biolabs), and those releasing DNA fragments of expected sizes (~2,900 and 1,200 bp) were sequenced using primers annealing close to the ends of the kanamycin resistance cassette (5′-TACTTTGCAGGGCTTCCCAACC and 5′-CGAAATGACCGACCAAGCGACG). Because strand transfer products were converted to fully double-stranded form before blunt-end ligation, the remainder of clones (21%) were discarded as cloning artefacts. Of all sequenced clones 99.3% contained pairs of donor DNA fragments joined to tDNA. Only unique clones from each *E. coli* transformation were used in further analyses. Deletions of various sizes (8–920 bp) accounting for 9.3% of all clones are explained by multiple integration events into a single target plasmid. The expected tDNA sequence duplication size of 4 bp was observed in the majority of remaining clones (98.5%), and only these were used in final sequence alignments (65 clones for WT, 49 for R329S, 45 for R329E and 58 for A188S). Sequence logos were generated using WebLogo[26].

21. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
22. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
23. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
24. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
25. Cherepanov, P. P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158,** 9–14 (1995).
26. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).

LETTER

# Iron–catalysed oxidation intermediates captured in a DNA repair dioxygenase

Chengqi Yi[1], Guifang Jia[1], Guanhua Hou[2], Qing Dai[3], Wen Zhang[1], Guanqun Zheng[1], Xing Jian[1], Cai-Guang Yang[1,4], Qiang Cui[2] & Chuan He[1]

Mononuclear iron-containing oxygenases conduct a diverse variety of oxidation functions in biology[1,2], including the oxidative demethylation of methylated nucleic acids and histones[3,4]. *Escherichia coli* AlkB is the first such enzyme that was discovered to repair methylated nucleic acids[5,6], which are otherwise cytotoxic and/or mutagenic. AlkB human homologues are known to play pivotal roles in various processes[7–11]. Here we present structural characterization of oxidation intermediates for these demethylases. Using a chemical cross-linking strategy[12,13], complexes of AlkB–double stranded DNA (dsDNA) containing 1,N^6-etheno adenine (εA), N^3-methyl thymine (3-meT) and N^3-methyl cytosine (3-meC) are stabilized and crystallized, respectively. Exposing these crystals, grown under anaerobic conditions containing iron(II) and α-ketoglutarate (αKG), to dioxygen initiates oxidation *in crystallo*. Glycol (from εA) and hemiaminal (from 3-meT) intermediates are captured; a zwitterionic intermediate (from 3-meC) is also proposed, based on crystallographic observations and computational analysis. The observation of these unprecedented intermediates provides direct support for the oxidative demethylation mechanism for these demethylases. This study also depicts a general mechanistic view of how a methyl group is oxidatively removed from different biological substrates.

E. coli AlkB is a prototype of iron-containing dioxygenases that catalyse oxidative demethylation of nucleic acids and histones[3,4]. Its human homologue ABH1 (also known as ALKBH1) demethylates 3-meC in DNA and RNA[7]; ABH2 (ALKBH2) guards the mammalian genome against N^1-methyl adenine (1-meA) damage[8]; ABH3 (ALKBH3) may repair methylated RNA damage[9,10]; and FTO, which demethylates 3-meT and N^3-methyl uracil (3-meU)[14,15], is a key factor in regulating energy homeostasis and obesity[11]. In addition, within this iron(II)/αKG-dependent subfamily, the JHDM proteins are engaged in human epigenetic regulation by catalysing the oxidative demethylation of methylated histones[4]. AlkB is quite versatile in that it can recognize and process three different types of base damage (Fig. 1 and Supplementary Fig. 1): (1) 3-meC and 1-meA, which formally bear a positive charge on the damaged nitrogen under physiological pH and are the most efficient substrates for AlkB[5,6]; (2) neutral 3-meT and N^1-methylated guanine (1-meG)[16–18]; and (3) cyclic adducts such as εA and 3,N^4-etheno cytosine (Supplementary Fig. 2)[19,20]. Using a previously developed disulphide cross-linking approach[12,13], we crystallized AlkB–dsDNA complexes containing 3-meC, 3-meT and εA, and solved the structures to high resolutions (Supplementary Table 1 and Supplementary Fig. 3). In these structures, the catalytically essential iron(II) was replaced by manganese(II), which occupies the same binding site, but does not support catalysis[21,22].

Comparisons of these structures, including a previously published 1-meA-containing structure (Protein Data Bank code: 3BIE), provide hints on how AlkB recognizes a diverse range of substrates with different hydrogen-bonding capacities and base dimensions. Of the four damaged bases, only one direct hydrogen bond is observed, which is

between the side chain of Asp 135 and the exocyclic amine of 1-meA or 3-meC (Supplementary Fig. 4). Mutation of this residue yielded mutant proteins (Asp 135 Ala, Asp 135 Ser or Asp 135 Asn) that display approximately 5–15% wild-type activity towards 1-meA at pH 7.0 (Supplementary Fig. 5)[23]. Analysis of natural bond orbital charge suggests that the exocyclic amine bears a large portion of the delocalized positive charge for 1-meA and 3-meC (Supplementary Fig. 6), which agrees with the observation that charged lesions in category (1) are preferentially recognized by AlkB. For the neutral lesions (3-meT and εA) that instead possess a hydrogen bond acceptor at the equivalent position, such an interaction is replaced by water-mediated hydrogen bonds (Supplementary Fig. 4). On the contrary, 3-meT is a substrate of FTO, and a hydrogen-bonding interaction between O4 of 3-meT and the amide nitrogen of Glu 234 has been observed in the crystal structure of FTO[24]; the disruption of this hydrogen bond results in the loss
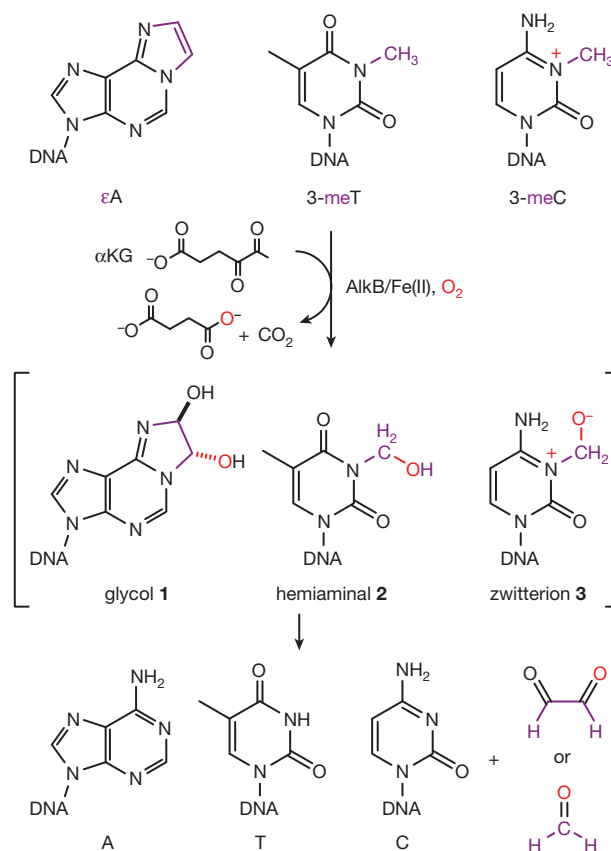
Figure 1 | Oxidative repair of damaged nucleic acid bases by AlkB.
Oxidative repair of εA, 3-meT and 3-meC by AlkB with intermediates glycol **1**, hemiaminal **2** and zwitterion **3** proposed in this study.

[1]Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. [2]Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706, USA. [3]Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. [4]Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China.

of the demethylation activity of FTO towards 3-meT. In the εA-containing structure, the cyclic adduct portion stacks against the side chain of Asp 133, which perhaps contributes to extra binding affinity of the lesioned base to the active site as compared to the repair product adenine (Supplementary Fig. 4a). On the other hand, substrates similar in size overlap well when bound in the active site, regardless of their hydrogen-bonding capacities (Supplementary Fig. 7). The final positions of a damaged purine and pyrimidine are consistently different, and are adjusted by AlkB so that the aberrant alkyl group is positioned towards the metal site for efficient catalysis (Supplementary Fig. 8 and Supplementary Note).

With several substrates covalently locked in the active site of AlkB, we set out to perform oxidation reactions in single crystals with a catalytically active iron centre (Supplementary Fig. 1). Performing enzymatic reactions in protein single crystals has been shown to be very effective in trapping and identifying unstable intermediates as well as elucidating reaction mechanisms[25–27]. The enzyme active sites provide nanoscale reaction vessels that allow isolation of otherwise transient intermediates.

Covalently linked AlkB–dsDNA complexes were crystallized with cofactors iron(II) and αKG under anaerobic conditions. Then, these crystals were exposed to air to initiate oxidation *in crystallo* (Supplementary Fig. 1). With the εA-containing DNA, a (S,S)-22,23-glycol intermediate **1** was captured (Fig. 2a and Supplementary Fig. 9a). The structure shows two newly formed hydroxyl groups that are involved in multiple hydrogen bonds: O26 forms a water-mediated interaction with Glu 136, and O24 hydrogen bonds directly to Arg 210 and succinate derived from αKG. The N6 atom of **1** forms indirect hydrogen bonds with the side chain of Asp 135 and the backbone amide of Lys 134, as also observed in the εA structure. Because O26 points away from the metal whereas O24 is within coordination distance to the iron (2.1 Å), we assigned O24 as the oxygen atom derived from O_2 oxidation of εA to yield the epoxide intermediate, which then undergoes ring opening by a water molecule attacking at C23 to afford the trans-glycol **1** (Fig. 2a and Supplementary Fig. 10a). The observation of this glycol intermediate firmly confirms the epoxidation repair mechanism of εA by AlkB. Subsequent protonation at N6 of **1** is anticipated to break N1–C22 and N6–C23 bonds and eventually produce an intact adenine base (Supplementary Fig. 10a). Indeed, species with molecular mass corresponding to epoxide and glycol **1** were observed by mass

spectrometry analysis[20]. Despite considerable efforts, we have yet to capture the epoxide intermediate in these crystals; the epoxide may not be as stable as **1** inside the enzymatic pocket of AlkB.

When 3-meT-containing crystals were exposed to air, the aberrant methyl group was oxidized to a hydroxymethyl group, which points towards the metal (3.6 Å) and hydrogen bonds to Arg 210, Asn 120 and succinate (Fig. 2b and Supplementary Fig. 9b). To our knowledge, this hemiaminal intermediate **2** is the first species of this type ever observed in oxidative demethylation in biology. The mono-dentate succinate shifts away from the newly formed hydroxymethyl group in this hemiaminal structure compared with the glycol structure, which may reflect different recognition of these intermediates in the enzymatic pocket (Supplementary Fig. 11).

We next studied *in crystallo* oxidation of 3-meC, a preferred substrate of AlkB. Exposing crystals that contain 3-meC with air led to observation of extra electron density on the aberrant methyl group (Supplementary Fig. 12). The best fitting model was obtained after releasing certain geometry restraints on angles and bond lengths of the hydroxymethyl group in the 3-hydroxymethylcytosine (3-hmC) dictionary file during refinement (Supplementary Table 2). In this model **3**, C20–O21 distance was refined to ~1.2 Å, which is shorter than a typical C–O single bond (~1.4 Å) (Supplementary Fig. 13a). Although protein crystallography at this resolution cannot conclusively assign the exact structure of **3** (Supplementary Fig. 13b), we chemically synthesized a 3-meC analogue, 3-deaza-3-methylcytosine (3-deazameC), to investigate further the 3-meC demethylation process (Supplementary Fig. 14). Crystals of AlkB–dsDNA containing 3-deazameC were also exposed to air, and in the resulting structure, an oxidized 3-deaza-3-hydroxymethylcytosine **4** (3-deazahmC) was unambiguously observed (Fig. 3a and Supplementary Fig. 15). Owing to C3 substitution, 3-deazameC is neutral and its oxidized product **4** is a stable alcohol. We also demonstrate that 3-deazameC can be converted to **4** by AlkB *in vitro*, although with a lower activity compared with 3-meC (Supplementary Fig. 16), as expected chemically.

We then subjected the crystal structure containing **4** to combined quantum mechanical/molecular mechanical minimizations (Supplementary Note). With a hydroxide ligand to iron(II) and a hydroxymethyl state of **4**, the optimized structure is stable and overlaps well with the crystal structure (Fig. 3b and Supplementary Fig. 17). Replacing the C3 carbon in the optimized 3-deazahmC structure to a nitrogen atom causes spontaneous proton transfer from the hydroxymethyl group of **4**
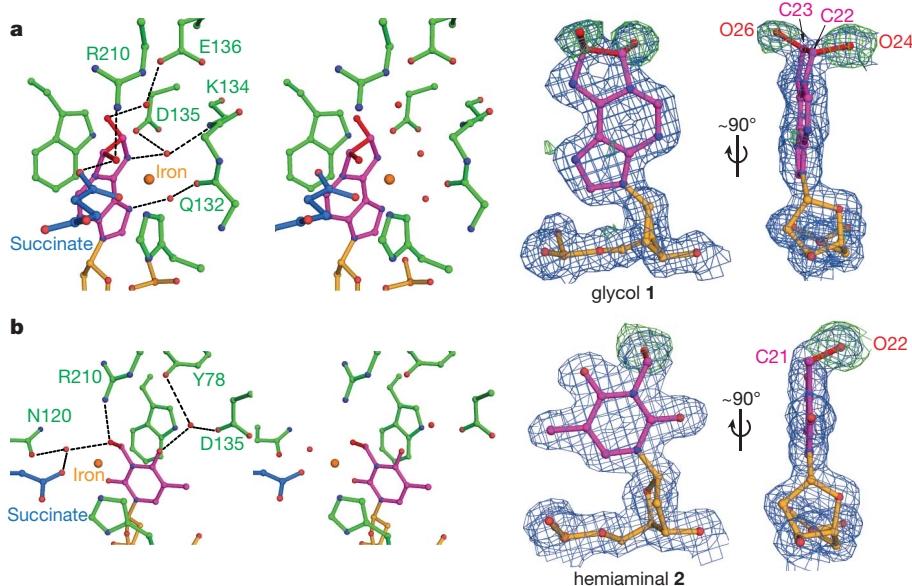
Figure 2 | Intermediates trapped during *in crystallo* oxidation of εA and 3-meT. **a**, Stereo views and electron density maps of the glycol intermediate **1** during εA repair. **b**, Stereo pairs and density maps of the hemiaminal intermediate **2** during the oxidation demethylation of 3-meT. The blue

$2F_{obs} − F_{cal}$ maps are contoured at 1.0σ and the green $F_{obs} − F_{cal}$ simulated-annealing omit maps were computed by removing extra atoms of intermediates (compared with the original substrates) and are contoured at 3.0σ. The extra atoms are shown in red. Hydrogen bonds are shown as dashed lines.
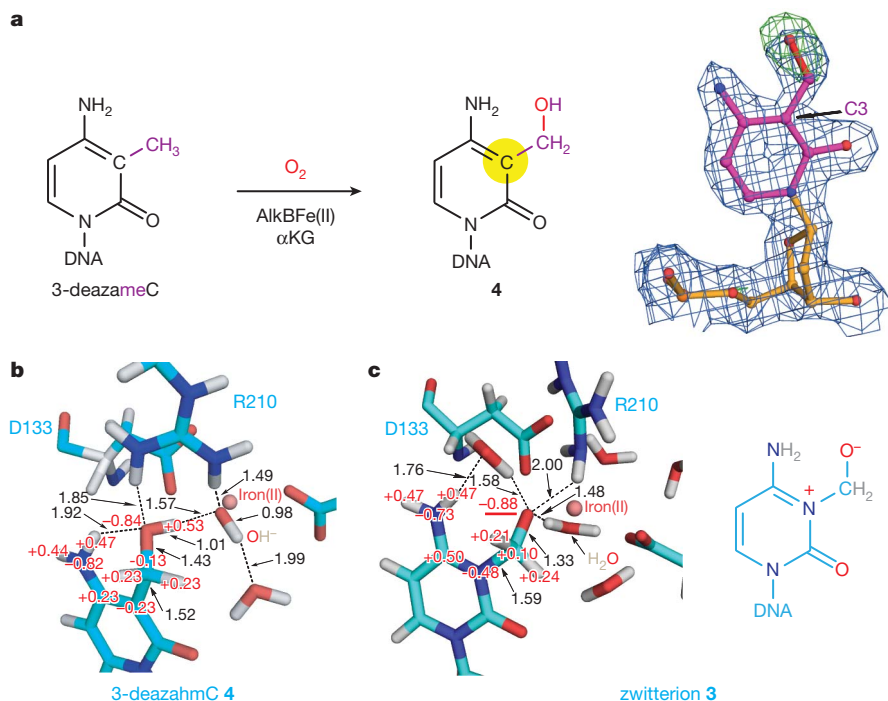
Figure 3 | A zwitterionic intermediate 3 is proposed for the demethylation of 3-meC. a, Oxidation of 3-deazameC in single crystals to yield 3-deazahmC, intermediate **4**. The density map and labels are generated and shown as in Fig. 2. **b**, Quantum mechanical/molecular mechanical calculated structure of **4**. **c**, Optimized structure of model **3** from the oxidized 3-meC crystal. For both calculated structures, carbon atoms are coloured in cyan, nitrogen in blue, oxygen in red, iron in pink and hydrogen in grey. Red natural bond orbital charges are labelled for several base atoms and key distances are marked in black (in ångströms).

to the hydroxide ligand, resulting in an iron(II)-H$_2$O state and a zwitterionic structure **5** (Supplementary Fig. 18a). In addition, an independent quantum mechanical/molecular mechanical minimization starting from the crystal structure with oxidized 3-meC **3** also leads to an iron(II)-H$_2$O state, and the optimized structure agrees with the crystal structure (Fig. 3c and Supplementary Fig. 18b). Analysis of the natural bond orbitals of the optimized structure containing **3** indicates that O21 bears a large negative partial charge, making the hemiaminal **3** actually a zwitterion (Fig. 3c), which is virtually the same as the computed **5**. Moreover, starting from the optimized model of **3**, replacing N3 with a carbon atom yields a converged structure that is very similar to the optimized structure containing 3-deazahmC (**4**), with the key features—iron(II)-OH$^-$ and the hydroxymethyl state—reproduced (Supplementary Figs 18c and 19). This difference in pK_a reflects the intrinsic chemical property of 3-meC, in which the cytosine ring bears a formal positive charge (Supplementary Fig. 20). Taken together, the evidence indicates that such a zwitterionic hemiaminal **3**, which explains the shortened C20–O21$^-$ distance, is very likely an intermediate of 3-meC oxidative demethylation.

E. coli AlkB and its human homologues ABH2 and ABH3 prefer positively charged 1-meA and 3-meC to neutral 3-meT and 1-meG as substrates. With neutral 3-meT, the intermediate **2** is a relatively 'stable' hemiaminal compared with that derived from 3-meC (Supplementary Fig. 10). Protonation at the O4 atom of **2** initiates bond migration and decomposition of **2** to liberate formaldehyde and yield the final intact thymine base (Supplementary Fig. 10b). In contrast, hydroxylation of 3-meC yields 3-hmC, which readily deprotonates to form the more stable zwitterionic intermediate **3** (Supplementary Fig. 10c). Because the positively charged cytosine base is a much better leaving group than the neutral thymine at physiological pH, the collapse of zwitterion **3** to cytosine and formaldehyde is expected to have a lower energetic barrier. This may partly explain the much faster repair of 1-meA and 3-meC by AlkB compared with 1-meG and 3-meT.

The charge-bearing feature of 1-meA and 3-meC is reminiscent of a positively charged trimethyl-lysine residue, the demethylation process

of which and its dedicated enzymes are of great interest (Supplementary Fig. 21)[4]. Within the AlkB family, FTO exhibits demethylation activity of neutral 3-meT and 3-meU, but not with 1-meA or 3-meC[14,15]. Depending on the charge state of the substrate, intermediates similar to **2** or **3** could form during the demethylation processes catalysed by these enzymes. The chemical nature of these intermediates can profoundly affect the reaction mechanism, reaction rate and substrate specificity (Supplementary Fig. 21).

In summary, AlkB has the ability to work on a diverse range of substrates. Besides the DNA repair function that maintains the integrity of the genome[28–30], members from this protein family also play diverse roles in biology[9–11]. The capture and structural characterization of several different intermediates presented in this study dissect the differences in the mechanism used by these enzymes to remove a methyl group oxidatively from biological substrates. This work also serves as an example of *in crystallo* reaction that leads to the trapping and characterization of otherwise unstable intermediates to help fully elucidate the reaction mechanism.

METHODS SUMMARY

The disulphide cross-linked protein–DNA complexes were purified using Mono-Q anion exchange chromatography, and crystals were grown using hanging-drop vapour diffusion methods. Diffraction data were collected from cryo-preserved crystals at beamlines 23ID-B, 21ID-D, 19BM-D and 14BM-C at the Advanced Photon Source, Argonne National Laboratory. The structures of AlkB–dsDNA complexes were solved by molecular replacement. Data collection and refinement parameters for all structures are given in Supplementary Table 1. Detailed procedures are presented in Supplementary Information.

1. Kovaleva, E. G. & Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nature Chem. Biol.* **4**, 186–193 (2008).

2. Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9,** 722–731 (1999).
3. Yi, C., Yang, C. G. & He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* **42,** 530–541 (2009).
4. Klose, R. J. & Zhang, Y. Regulation of histone methylation by demethylimination and demethylation. *Nature Rev. Mol. Cell Biol.* **8,** 307–318 (2007).
5. Falnes, P. O., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli. Nature* **419,** 178–182 (2002).
6. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419,** 174–178 (2002).
7. Westbye, M. P. *et al.* Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* **283,** 25046–25056 (2008).
8. Ringvoll, J. *et al.* Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO J.* **25,** 2189–2198 (2006).
9. Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421,** 859–863 (2003).
10. Sundheim, O. *et al.* Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J.* **25,** 3389–3397 (2006).
11. Frayling, T. M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316,** 889–894 (2007).
12. Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452,** 961–965 (2008).
13. Qi, Y. *et al.* Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* **462,** 762–766 (2009).
14. Gerken, T. *et al.* The obesity-associated *FTO* gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318,** 1469–1472 (2007).
15. Jia, G. *et al.* Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.* **582,** 3313–3319 (2008).
16. Koivisto, P., Robins, P., Lindahl, T. & Sedgwick, B. Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.* **279,** 40470–40474 (2004).
17. Falnes, P. O. Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.* **32,** 6260–6267 (2004).
18. Delaney, J. C. & Essigmann, J. M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli. Proc. Natl Acad. Sci. USA* **101,** 14051–14056 (2004).
19. Mishina, Y., Yang, C. G. & He, C. Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins. *J. Am. Chem. Soc.* **127,** 14594–14595 (2005).
20. Delaney, J. C. *et al.* AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo. Nature Struct. Mol. Biol.* **12,** 855–860 (2005).
21. Yu, B. *et al.* Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439,** 879–884 (2006).
22. Yu, B. & Hunt, J. F. Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc. Natl Acad. Sci. USA* **106,** 14315–14320 (2009).
23. Holland, P. J & Hollis, T. Structural and mutational analysis of *Escherichia coli* AlkB provides insight into substrate specificity and DNA damage searching. *PLoS ONE* **5,** e8680 (2010).
24. Han, Z. *et al.* Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464,** 1205–1209 (2010).
25. Schlichting, I. *et al.* The catalytic pathway of cytochrome p450cam at atomic resolution. *Science* **287,** 1615–1622 (2000).
26. Kovaleva, E. G. & Lipscomb, J. D. Crystal structures of Fe^{2+} dioxygenase superoxo, alkylperoxo, and bound product intermediates. *Science* **316,** 453–457 (2007).
27. Burzlaff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401,** 721–724 (1999).
28. David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447,** 941–950 (2007).
29. Hitomi, K., Iwai, S. & Tainer, J. A. The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair* **6,** 410–428 (2007).
30. Fromme, J. C., Banerjee, A. & Verdine, G. L. DNA glycosylase recognition and catalysis. *Curr. Opin. Struct. Biol.* **14,** 43–49 (2004).

METHODS

Oligonucleotide synthesis. Oligonucleotides containing a disulphide-tethered cytosine and damaged bases (εA, 3-meT and 3-meC) were prepared by incorporating the O⁴-triazolyl-dU-CE Phosphoramidite (Glen Research), etheno adenosine phosphoramidite, N³-methyl thymidine CED phosphoramidite and N³-methyl deoxy cytidine CED phosphoramidite (the last three were purchased from ChemGenes) at the desired position during solid-phase synthesis[31]. All synthetic oligonucleotides were purified with reverse-phase high-performance liquid chromatography.

Cross-linking and purification of the AlkB–dsDNA complexes. A truncated AlkB with deletion of the amino (N)-terminal 11 amino acids was cloned into a pET30a vector (Novagen) and overexpressed in *E. coli* BL21(DE3)[21]. The protein was purified following a previously described procedure[32]. The Ser 129 to Cys mutation was introduced using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Synthetic oligonucleotides (5′-TAGGTAAXAC*CGT, where C* is a disulphide-tethered cytosine and X stands for εA, 3-meT or 3-meC, and its complement strand 5′-AACGGTT_TACCT-3′, where '_' stands for an abasic site) were prepared, annealed and cross-linked to the S129C mutant AlkB as described[12]. The covalently linked AlkB–dsDNA complexes were purified using Mono-Q anion exchange chromatography (GE Healthcare), after which the buffer was exchanged to 100 mM NaCl, 10 mM Tris-HCl (pH 7.4), and the complexes were concentrated to 12–15 mg ml⁻¹ before drops were set up.

Crystallization of the AlkB–DNA complexes containing Mn(II) and αKG. For the crystals grown under aerobic conditions, 1.0 mM MnCl₂ (Mn(II) is able to occupy the Fe(II) site, but is incapable of catalysis) and 2.0 mM αKG were added to the complexes; they were mixed in a 1:1 ratio with well solution containing 100 mM NaCl, 25 mM MgCl₂, 100 mM cacodylate (pH 6.5) and 20–24% w/v PEG 8K or 4K. Crystals appeared after 2–4 days and were allowed to grow for several weeks. They were transferred into a cryo-protectant solution containing the reservoir solution plus 20% glycerol, and frozen in liquid nitrogen for X-ray data collection.

Conditions for *in crystallo* oxidation reactions. All buffers and solutions used in crystallization trials were degassed with N₂ for 30 min, and then transferred into an anaerobic chamber to equilibrate for at least 16 h before use. The purified AlkB–dsDNA complexes were buffer-exchanged with degassed buffer at least four times, and then equilibrated in the anaerobic chamber before setting up crystallization trials. To set up drops, 1.0 mM (NH₄)₂Fe(SO₄) and 2.0 mM αKG (the 50 mM stock solutions of which were always freshly prepared by dissolving (NH₄)₂Fe(SO₄) and αKG solid in pre-equilibrated Milli-Q water in the anaerobic chamber) were added to the complexes (around 0.35–0.45 mM), and they were mixed in a 1:1 ratio with well solution containing 100 mM NaCl, 50 mM MgCl₂, 100 mM cacodylate (pH 6.5) and 20–24% w/v PEG 8K or 4K. Crystals appeared after 2–4 days, after which they were exposed to air by opening the cover slides periodically and then immediately frozen in liquid N₂ for X-ray data collection. Different time points of air exposure were sampled (2–4 min, 1–2 h, 5–6 h, 2 d and 9 d). Because these crystals are sensitive to oxygen, drops were normally set up no more than 2 weeks in advance from the actual date of X-ray data collection.

Air exposure of the AlkB–DNA complex containing 3-meC, Mn(II) and αKG. Mn(II) has previously been shown to occupy the metal-binding site of AlkB, but is catalytically inactive[21,22]. To be particularly cautious, the *in crystallo* oxidation procedure for the Fe(II)/αKG-containing AlkB–dsDNA complex was performed for the Mn(II)/αKG-containing AlkB–dsDNA complex. The base lesion 3-meC was selected because it was the most efficient substrate of AlkB among the three damaged bases tested (εA, 3-meT and 3-meC). Crystallization drops were set up under anaerobic conditions and cover slides were opened to expose these Mn(II)/αKG-containing crystals to air. Two independent batches of such crystals were tested and crystals were picked up after exposure to air for 2 h and 6 h, respectively. The respective structures of these crystals were solved to 1.92 Å and 1.98 Å, with R_{work}/R_{free} of 17.9/22.5 and 18.9/24.3. The αKG cofactor has good electron density in both structures, indicating that Mn(II) is incapable of oxidation catalysis; the 3-meC base remains unmodified in both structures as well (Supplementary Fig. 22). Because these structures are the same as the structure 3O1M ((manganese/αKG)AlkB-DNA: 3-meC), they were not deposited in the Protein Data Bank database. In summary, these control experiments confirm AlkB-mediated *in crystallo* oxidation with Fe(II) as the catalytically essential metal ion.

Structure determination and refinement. The AlkB–DNA complex structure was phased by molecular replacement (using Phaser)[33], using the previously published AlkB structure as a search model. The model was built by using COOT and refined with the program REFMAC5 from the CCP4 suite[34,35]. Simulated annealing omit maps were generated with Phenix[36]. Data collection and refinement parameters for all structures are given in Supplementary Table 1. Molecular graphics figures were prepared with PyMOL[37].

Ligand refinement and molecular modelling. Regular damaged bases (εA, 3-meT and 3-meC) were used as the initial model for the crystals obtained from *in crystallo* oxidation. Based on the resultant $F_{obs} – F_{calc}$ maps, appropriate modifications were modelled into the density to obtain the best fit of the data. Dictionaries for the glycol and hemiaminal intermediates were produced with Monomer Library Sketcher (in CCP4 suite). Occupancy of the additional atoms of the intermediate structures was approximately 100% based on the average temperature factors (which are similar to those of the rest of the base atoms) and the lack of $F_{obs} – F_{calc}$ density contoured at 3.0σ. Model bias associated with ligand refinement was evaluated by, in addition to simulated annealing omit maps using Phenix, examining omit difference maps. The extra atoms were removed from the final refined models, and resulting models were allowed to refine with at least ten cycles of restrained refinement. The resultant positive density in the difference $F_{obs} – F_{calc}$ maps was found to agree with the refined positions and structures of the base intermediates in the full models.

Reproducibility of the observed intermediates. The three intermediate structures reported here were also observed in several additional, independently refined structural data sets obtained from crystals prepared using the same procedure described (data not shown). These observations affirm a physical basis for the observed reactivity in crystalline form.

Synthetic protocols of compound XIII from IV. Compound **IV** was obtained according to literature methods and was directly isolated from the α-anomer by careful column chromatography[38].

4-Amino-1-(2′-deoxy-β-D-erythro-pentofanosyl)-2-pyridone (X). To a solution of **IV** (200 mg, 0.418 mmol) in methanol (5 ml) was added 25% sodium methoxide (0.5 ml) dropwise; the resulting solution was stirred at room temperature for 30 min. Acetic acid was added dropwise to adjust the pH to 7. The solvents were removed under reduced pressure and the residue was purified by silica gel chromatography, eluting with 5–15% MeOH in CH₂Cl₂, to give **X** (90.4 mg, 90%) as white foam. ¹H NMR (500.1 MHz) (CD₃OD) δ: 7.64 (d, J = 7.5 Hz, 1H), 6.51 (t, J = 6.5 Hz, 1H), 6.02 (d, J = 7.5 Hz, 1H), 4.36 (m, 1H), 3.94 (m, 1H), 3.80 (dd, J = 12.0, 4.0 Hz, 1H), 3.74 (dd, J = 12.0, 4.0 Hz, 1H), 2.37 (m, 1H), 2.10 (m, 1H), 1.91 (s, 3H). ¹³C NMR (125.8 MHz) (CD₃OD) δ: 162.8, 154.6, 130.0, 100.3, 87.1, 85.2, 70.8, 61.6, 41.0, 8.0.

4-Phenoxyacetylamido-1-(2′-deoxy-β-D-erythro-pentofanosyl)-2-pyridone (XI). To a solution of **X** (82 mg, 0.34 mmol) in pyridine was added trimethylsilyl chloride (10 eq., 0.425 ml); the mixture was stirred at room temperature for 2 h. Then phenoxyacetyl chloride (0.236 mL, 5 eq.) was added and the mixture was stirred for another 2 h. Ammonia hydroxide (1 ml) was added and stirred for 1 h. Removal of the solvents gave a residue, which was purified by silica gel chromatography, eluting with 3–10% MeOH in CH₂Cl₂, to give **XI** (100 mg, 78%) as white foam. ¹H NMR (500.1 MHz) (CD₃Cl+CD₃OD) δ: 7.88 (d, J = 8.0 Hz, 1H), 7.33 (m, 2H), 7.24 (d, J = 8.0 Hz, 1H), 7.03 (m, 1H), 6.97 (d, J = 8.5 Hz, 2H), 6.40 (t, J = 6.0 Hz, 1H), 4.66 (s, 2H), 4.43 (m, 1H), 3.97 (m, 1H), 3.81 (m, 1H), 3.74 (m, 1H), 2.46 (m, 1H), 2.11 (m, 1H), 2.01 (s, 3H). ¹³C NMR (125.8 MHz) (CD₃Cl+CD₃OD) δ: 171.6, 166.7, 160.8, 147.7, 134.3, 133.7, 126.4, 118.5, 106.2, 91.4, 90.4, 74.3, 71.3, 65.4, 45.3, 13.5.

4-Phenoxyacetylamido-1-((5′-O-(4,4′-dimethoxytrityl)-2′-deoxy-β-D-erythro-pentofanosyl)-2-pyridone (XII). Ninety milligrams (0.24 mmol) of **XI** was dissolved in pyridine (3 ml), and 4,4′-dimethoxytrityl chloride (97.6 mg, 0.288 mmol) was added while stirring the solution. After being stirred overnight at room temperature, the reaction mixture was quenched with MeOH (1 ml) and stirred for an additional 5 min. The reaction mixture was concentrated to dryness under vacuum. Dichloromethane (100 ml) was added and washed with sodium hydrogen carbonate (5%, 50 ml) and brine and then dried over sodium sulphate. After the organic phase was concentrated to dryness, the residue was purified by silica gel chromatography, eluting with 1–3% MeOH in dichloromethane containing 0.2% Et₃N, to give **XII** (145 mg, 91%) as a white foam. ¹H NMR (500.1 MHz) (CD₃CN) δ: 7.71 (d, J = 7.5 Hz, 1H), 7.47 (d, J = 7.5 Hz, 2H), 7.34 (m, 9H), 7.25 (m, 1H), 7.05 (m, 2H), 6.92 (d, J = 6.5 Hz, 1H), 6.88 (m, 4H), 6.39 (m, 1H), 4.67 (s, 2H), 4.43 (m, 1H), 3.98 (m, 1H), 3.78 (m, 6H), 3.35 (m, 2H), 2.46 (m, 1H), 2.11 (m, 1H), 1.99 (s, 3H). ¹³C NMR (125.8 MHz) (CD₃CN) δ: 162.0, 158.7, 157.3, 149.7, 144.9, 143.2, 135.9, 135.8, 130.08, 130.06, 129.9, 129.8, 128.1, 128.0, 127.0, 123.8, 122.0, 117.3, 116.8, 114.8, 114.2, 113.1, 101.3, 86.4, 85.9, 85.5, 70.2, 67.3, 63.0, 54.9, 41.2, 31.5, 9.6.

4-Phenoxyacetylamido-1-[(5′-O-(4,4′-dimethoxytrityl)-3′-O-(2-cyanoethyl-N,N-diisopropyl)phosphoramidite-2′-deoxy-β-D-erythro-pentofanosyl]-2-pyridone (XIII). One hundred and forty milligrams (0.21 mmol) of **XII** was dissolved in dry CH₂Cl₂ (5 ml) and 1-methylimidazole (3.40 mg, 41.5 μmol). N,N-diisopropylethylamine (156 mg, 0.84 mmol) was added to the stirring solution followed by 2-cyanoethyl N,N-(diisopropylchloro)-phosphoramidite (100 mg, 0.42 mmol). After being stirred at room temperature for 2 h, the reaction mixture was added to dichloromethane (5 ml) and the mixture was washed with 5% aqueous sodium bicarbonate and brine, dried over sodium sulphate and concentrated.

The residue was purified by silica gel chromatography, eluting with 10–12% acetone in dichloromethane containing 0.2% Et_3N, to give **XIII** (150 mg, 81%) as a white foam. ^{31}P NMR (202.5 MHz) (CD_3CN) δ: 148.3 and 148.4 p.p.m. by high-resolution mass spectrometry for $C_{49}H_{57}N_4NaO_9P$, $[MNa]^+$ 399.3761 (calculated); 899.3753 (found).

31. Mishina, Y. & He, C. Probing the structure and function of the *Escherichia coli* DNA alkylation repair AlkB protein through chemical cross-linking. *J. Am. Chem. Soc.* **125,** 8730–8731 (2003).
32. Mishina, Y., Chen, L. X. & He, C. Preparation and characterization of the native iron(II)-containing DNA repair AlkB protein directly from *Escherichia coli. J. Am. Chem. Soc.* **126,** 16930–16936 (2004).
33. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57,** 1373–1382 (2001).
34. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
35. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
36. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
37. DeLano, W. L. The PyMOL molecular graphics system. <http://www.pymol.org> (2002).
38. Searls, T. & McLaughlin, L. W. Synthesis of the analogue nucleoside 3-deaza-2'-deoxycytidine and its template activity with DNA polymerase. *Tetrahedron* **55,** 11985–11996 (1999).

CORRECTIONS & AMENDMENTS

Global threats to human water security and river biodiversity

C. J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon,
A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan,
C. Reidy Liermann & P. M. Davies

In this Article, the full present address for author P. B. McIntyre was inadvertently missing from the bottom of the page. The correct present address is: Center for Limnology, University of Wisconsin, Madison, Wisconsin 53706, USA. This has been corrected in the online PDF.

CAREERS

| |

AWARDS

Thumbs up for Canadian mentors

Nature *recognizes the best of the country's science advisers.*

BY NICOLA JONES

John Smol is one of Canada's top researchers in environmental science, his studies ranging from acid rain and climate change to the effects of pollution on salmon populations and the Arctic — yet he always has time for his students. Chris Wood's students marvel at how he infuses others with his work ethic and enthusiasm — almost, they say, as if through "subliminal suggestion". Now, both mentors are receiving recognition for their abilities to guide and inspire.

Smol and Wood are the winners of this year's *Nature* awards for mentorship (see 'Distinguished careers'). Smol, a palaeoecologist at Queen's University in Kingston, Ontario, won for mid-career mentoring achievement. Wood, an expert in fish physiology at McMaster University in Hamilton, Ontario, won for lifetime

achievement. Each received Can$10,000 (US$9,800) on 5 November, at a ceremony in Toronto, Ontario.

The winners were chosen from more than 50 nominees, put forward in applications by colleagues, former students and the nominees themselves. The judging committee was led by John Hepburn, research and international vice-president at the University of British Columbia, Vancouver. The panel remarked on the high quality of all the applications, but was won over by the passionately sincere testimonials from Wood's former students, many of whom spoke of his lab as if it were a family. And Smol is in a class of his own, says Hepburn. "You get the whole package," he adds. "You get the picture of a totally socially engaged scientist."

This is the seventh year that *Nature* has sponsored an award to recognize excellence in mentorship, each time focusing on a different country.

MARINER OF HUMAN KINDNESS

Smol, 54, founder and co-director of the Paleo-ecological Environmental Assessment and Research Lab at Queen's University, has earned the nickname 'mariner of human kindness'. He says he gets as much out of interacting with his students as they do, thanks to their probing questions. "We have friendly jousting," he says. "I think that's how they learn. It's how I learn too." The dedication in his most recently co-authored ecology textbook reads: "To my students, who have inspired me far more than I could ever have inspired them."

Hepburn says that Smol's most striking accomplishment is his success rate. "He has never failed with a graduate student — never," he adds. The application lists more than 60 graduate students who have passed through Smol's lab, none of whom have withdrawn or failed. "That's mentoring," says Hepburn. Smol attributes this success to customizing his approach. "Part of the job is not just to be a science adviser, but to help people through whatever happens, from deaths to personal-relationship problems," he says. This included supporting one student when he blew the whistle on a plagiarist, and keeping another's research programme alive when she could not work for a year because of health and personal problems. "I went to bat for a lot of my students," he says.

The judges praised Smol for emphasizing the mentor's role in training students not only in data collection, but also in communication and ethics, ensuring that they come out with a "strong moral compass", as Smol put it in his ▶

application. "We should never underestimate the importance of what we do as scientists," he says, noting the value of, for example, pressing politicians to create evidence-based policy. "If your data is strong, you have to make the case. It's your job," says Smol. He spends a lot of time talking to the media and policy-makers, and hopes in this way to lead by example.

Wood, 63, has made his name in the field of comparative physiology, spearheading studies of fish responses to pollution (including acid rain) and climate change that have made an impact on public policy in Canada. The former students who nominated him all commented on his infectious energy and passion for science. They praised his ability to inspire and impress others with his own work ethic — that mentoring power of subliminal suggestion. More seriously, they call it "mentorship through example". "I've never really understood why my students have done quite so well, but I'm enthusiastic and I think that comes across," says Wood.

Wood has attracted almost 100 postdocs and graduate students to his lab; about 40% now hold faculty positions. "Nearly single-handedly, he is repopulating the universities of the world with the next generation of comparative physiologists," wrote nominator Steve Perry, a former postdoc from Wood's lab and now vice-dean of research at the University of Ottawa in Ontario. Wood has also overseen more than 220 undergraduate students.

Two of Wood's student nominators have won the governor-general's gold medal, awarded to the top doctoral candidate at each university in Canada; one notes that Wood called his undergraduate grades "colourful" when they first met. "His marks were spotty — he had to repeat a year," says Wood. "But he

"The job is not just to be a science adviser, but to help people through whatever happens."

Canada's best mentors: John Smol (left) and Chris Wood inspire their students and postdocs.

did such a good job in his thesis that I made a pitch for him." After fighting to get the student into graduate school, Wood says the student exceeded his "wildest expectations".

Perhaps Wood's most powerful technique is holding compulsory weekly lab meetings. Many of his nominators said that the gatherings made them feel like part of a collaboration that was greater than the sum of its parts. "It's a bonding experience," says Wood. "My only rule is that you can't miss this meeting."

TWO OF A KIND

The winners were both complimented for their healthy and fun lab atmosphere, which keeps enthusiasm high and puts troubles into perspective. Smol says that he aims to laugh out loud with his students at least once a day — about anything from work to clothes to a television programme. Wood hosts an annual get-together, at which he once greeted his newest postdoc while dressed in drag as a member of the pop group the Spice Girls. "In our field we have this philosophy of having a good time — working hard, playing hard," says Wood. "It turns people on to the idea that science doesn't

have to be dry." Both mentors have had several successful female students in their labs.

Each says that getting students out of the lab and into conferences or visits to other research groups is also a vital part of the mentoring strategy. "John is ready to forgo hiring administrative help (despite pleas from his colleagues who fear for his health!) and instead uses his discretionary funding to send students to important learning and research events," wrote nominator Brian Cumming, a former graduate student of Smol's and now director of the School of Environmental Studies at Queen's University. Wood often takes a mixture of current and former students to the remote Bamfield Marine Sciences Centre on the west coast of Vancouver Island. "I think that's really important in forging intergenerational links," says Wood, who plans to give a chunk of his award money to the centre.

The unifying theme across all the entries for the awards, says judging-panel member John Clague, an Earth scientist at Simon Fraser University near Vancouver, is that none of the applicants sees their students as just another set of hands. "It isn't all about increasing productivity" as it can be in some labs, he says.

The judges say that there is a need for more recognition of mentoring in the university system. Some efforts are already under way. Many Canadian universities have an award for excellence in graduate supervision; Queen's University is one, and Smol was the inaugural winner in 2006. But the judges agreed that more could be done. "We're at the same place with mentorship now as we were with teaching ten years ago" in terms of official recognition, says Hepburn.

"We train scientists to do experiments," says judging-panel member Mona Nemer, vice-president of research at the University of Ottawa. "They come out as thinkers and problem solvers in their areas. But we don't train them in management, ethics, communication and teaching. That's where mentorship comes in." Both Smol and Wood are clearly carrying that torch. ∎

CASE STUDY
Distinguished careers

John Smol
- Canada Research Chair; fellow of the Royal Society of Canada; founder of the palaeoecology lab at Queen's University in Kingston, Ontario.
- Notable awards: 2004 Herzberg Gold Medal for science and engineering; 2006 Queen's University Award for Excellence in Graduate Supervision; 2009 3M National Teaching Fellowship.
- Size of lab: ~20 students.
- Years of supervision: ~25.
- Graduate students supervised: ~60.
- Postdocs: ~20.

Chris Wood
- Canada Research Chair; fellow of the Royal Society of Canada; professor of biology at McMaster University in Hamilton, Ontario.
- Notable awards: 1999 Fry Medal for Lifetime Achievement; 2002 Lifetime Achievement Award, American Fisheries Society; 2007 Miroslaw Romanowski Medal in environmental science.
- Size of lab: ~15 students.
- Years of supervision: ~35.
- Graduate students supervised: ~50.
- Postdocs: ~40.

Nicola Jones is a freelance writer and editor based in Pemberton, British Columbia.

TURNING POINT
Jacob Hanna

In April 2011, Jacob Hanna, currently a Genzyme postdoctoral fellow at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, will start his own lab at the Weizmann Institute of Science in Rehovot, Israel. As one of the few Palestinian researchers at the Weizmann, he is eager to build scientific bridges in the region.

Why did you choose to return to Israel to start your independent lab?
There were two major factors. The Weizmann is a great multidisciplinary centre that offers strong financial and organizational support. And on a personal level, as a Palestinian who grew up in Israel, it is meaningful to do science there.

Do you plan to encourage more scientific exchange between regions?
I am a scientist, and research is my priority. But building a diverse group of students and collaborators at the Weizmann is one way to have a positive impact. Palestinians are underrepresented in academia, in part because there is little academic infrastructure in the area outside Israel, and if I can encourage Israeli and Palestinian scientists to join my lab and work in an integrative environment, I will be happy.

You have focused on stem-cell research projects with a high risk of failure. Why?
The Whitehead has a risk-taking culture. My mentor, Rudolf Jaenisch, encourages us to take on high-risk ideas and not to be constrained by technical limitations. That support taught me to work hard and be adventurous. He also tells us not to lower standards by overlooking complexity. In this field, there is a tendency to rely on concepts that are not yet well defined. I'm realizing how important it is to do quantitative experiments to get conclusive answers.

Is media attention affecting stem-cell research?
It is a double-edged sword. I'm happy that the importance of stem cells has been recognized. But I'm concerned that media attention may be directing the science by encouraging scientists to publish too early. I worry that this rush lowers standards.

What was your best career decision?
I did an MD–PhD at the Hebrew University of Jerusalem. Towards the end of my studies, I realized that I enjoyed medicine and research, but would be better at research. So I focused on that instead of doing a residency. I'm glad I have my medical degree. It gives me a different view of research. For example, as a physician, I'm interested in type I autoimmune diabetes and genetic susceptibilities. With stem cells, we can ask questions about how to humanize mouse models of this disease to understand its genesis.
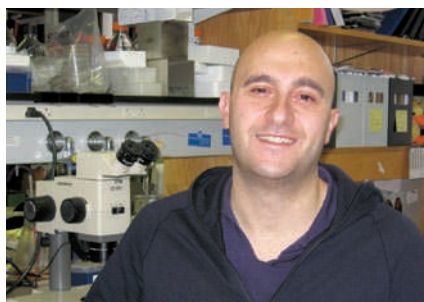
Any missed opportunities you regret?
No. In fact, I'm glad I passed on an early opportunity to start my own lab after one year of my postdoc. I was fortunate that my research went well in that first year, but I realized I would be rushing. I know it is time now, because I have a much better grasp of the important questions and I feel confident that I have gained enough experience and maturity to handle any scientific situation.

What questions will you ask at the Weizmann?
I want to understand the continuum of reprogramming cells to pluripotency. But I also want to develop a new avenue of research — disease-specific stem-cell research. I want to use technology to ask basic immunology questions and develop human models. My first challenge will be to forge collaborations throughout Israel, to lay the foundation for my future research.

What inspires your work?
Mostly it is the work of others. When I read somebody else's great paper, even in another field, I want to do something similar — come up with a new idea that will make a contribution. What drives me is good scientific work.

What is your career philosophy?
Work hard and never admire your own science. It can always be better. ∎

INTERVIEW BY VIRGINIA GEWIN

Website to help women

An interactive website aims to halt attrition among female doctoral students in science, technology, engineering and maths. Launched on 4 November by Arizona State University (ASU) in Tempe, CareerWISE (go.nature.com/nbs6qv) gives coaching and support to help women overcome hostile environments and balance priorities. Content includes videos of female scientists who have cleared graduate-school hurdles. The site seeks to help women develop coping strategies, says project leader Bianca Bernstein, an ASU psychologist. Her research on discrimination among female graduate students who felt unable to seek support from colleagues helped inspire the site. Two National Science Foundation grants totalling US$3.2 million funded the site.

EUROPE
Universities ranked

University College London (UCL) tops a list of biology programmes at 130 high-performing universities in Europe, says a report released on 27 October by the Centre for Higher Education Development (CHE), a non-profit group in Gütersloh, Germany. As part of an effort to promote educational reform, the CHE assessed universities' biology, chemistry and physics programmes. It measured the institutions' number of publications, citations and outstanding researchers; student and staff mobility; presence of Marie Curie fellows and other programmes; and whether researchers had a European Research Council grant or Nobel prize. UCL did well in all categories except mobility.

TENURE
Satisfaction survey

A Harvard University group has initiated a survey to gauge the satisfaction of tenured faculty members in relation to research, teaching and other activities. Harvard's Collaborative on Academic Careers in Higher Education (COACHE) in Cambridge, Massachusetts, began its pilot phase in October, following efforts to track the satisfaction of tenure-track staff. "Tenured faculty are key to institutional climate, the most important thing for tenure-track faculty," says COACHE director Kiernan Mathews. Topics will include recognition, work–life balance, retention policies and time for duties such as research and teaching. COACHE plans to launch the full survey by autumn 2011.

GEORGE AND PRITI

Binary love.

ANAND ODHAV NARANBHAI

Priti Sinvi, 59: "Oh, George."
George Kwame, 67: "Oh, Priti."
Priti, a Kenyan-born extremophile biologist doing postdoctoral work at the JS Moons Institute and recent recipient of the 'African Academy of Science Young Scientist Award for Stellar Research': "I love you, George."

George, the smartest engineering graduate on the continent, who discovered a new particle while on a summer elective on V645 Centauri and then, for his PhD, perfected the use of the particle as *the* next-generation fuel source: "I love you, Priti."

They nestled in a quiet, hidden corner of the spaceship's maintenance floor, a wall of glass separating them from the loneliness of space. Two bodies fused into one, a shadow against the white splendour of a solar flare.

"George, can it be?"

"Yes, Priti. You are my life-force, my reason for breathing. I can't live without you."

"Oh, George, I wish that were enough. But life is much more complicated than that. We have to think about the long term. You and I … we're so different. Not morally or habitually or like that, but culturally. You know what I'm talking about, George."

"Priti, I've told you not to listen to your parents too much. I suppose we will have differences like all couples do, but we can work through them. Our only difference is the colour of our skin. I love you too much to let outdated opinions based on an ancient pigment get in the way."

"You can say that easily now, George, but it will be very different once we have been together for a while and this dream that we are in senesces. And you know, my parents have been around for 200 years and both are on their third marriages, so they know how relationships work. [*Sighs*] I keep thinking about my twin aunts who ran away with handsome boys from Mars, only to end up in messy divorces. One of them is now an alcoholic theoretical dramatist and the other one is an artist-in-residence on Pluto!"

"It doesn't matter what happens to other people. I've told you, I don't care about that and I'm not like them. I love you for who you are on the inside and outside."

Priti put her head on George's chest. George held her in his gloved arms.

"Oh, George. I don't know what to do. My heart is telling me one thing, and my brain another. Whenever I am confused about a decision, I always listen to my brain. I always try to do what's best for survival, what makes the most evolutionary sense. I try to do the most rational thing, because reason is the only thing that you can trust when your mind is wrought with emotional complexity."
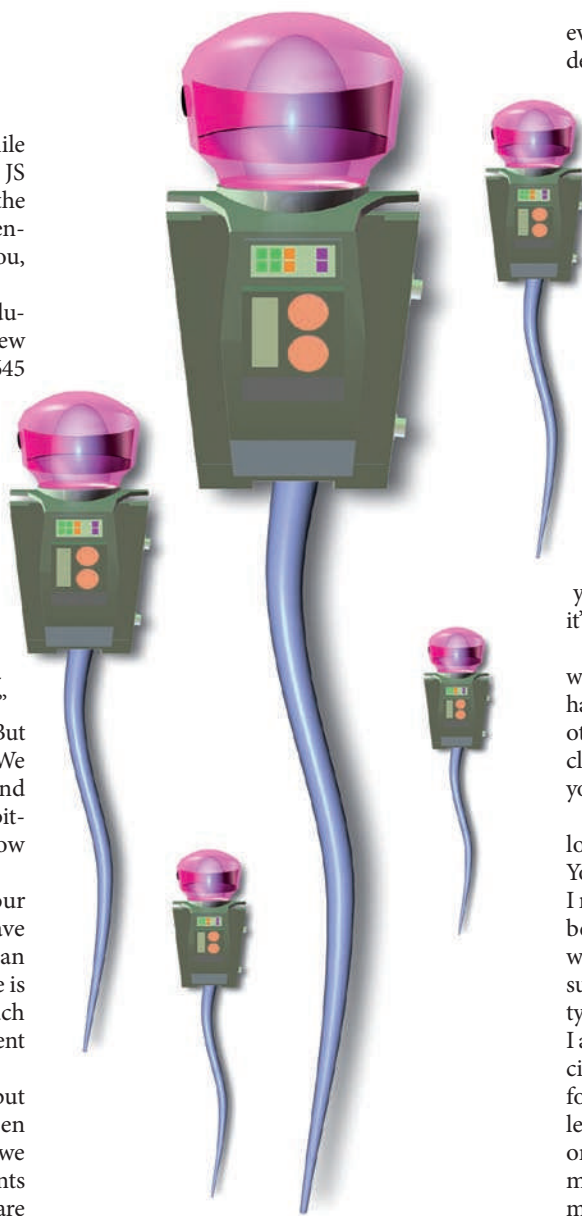
"Have you tried drawing those advantage and disadvantage tables you always use to help you decide?"

"I did, George. I spent days writing down everything and forced myself to go through every permutation. At the end, I just couldn't decide which column outweighed the other. Whenever I choose the second column, my heart just can't accept it. My subconscious neural circuits become restless."

"That means it's love, Priti."

"I know it's love, George. But love is an effervescent and evanescent thing."

"Not mine!"

"Oh, George! Don't try to charm me. We have to be serious. We live at opposite ends of the Solar System. We'll only see each other once, maybe twice every five years."

"I'm not trying to charm you, Priti. I'm being as honest as I can. We'll talk to each other every day and can see each other after work using the virtual conferencing function on that cellphone I gave you for your birthday. Even if we are light years apart, it'll be like we're right next to each other."

"It's not the same. There's so much more that words alone can't possibly express. George, we have to think carefully about this. There'll be other girls that you'll meet, you know. Girls closer than me. Girls that will look prettier, younger, more qualified than me."

"That's not fair. You know I don't even look twice at a girl with a short spacesuit. You're the only one I have eyes for, Priti. Hell, I never even thought about another woman before you. I just kept my head down and worked at my fuel cell. Priti, I may not be a suave, weight-pumping, grand-commander-type guy, but I know I am a good man, I know I am a smart man, I know I'm a man of principle, a man of science. I will take care of you for the next 200 million years. I will never let any water shortage or ship malfunction or stupid cancer take my life partner from me. I will never let anybody interfere with my girl's work out of professional jealousy. I will fly to the ends of the Universe without going to the loo once for you. I will suffer any pain, any punishment, any injury, any disappointment, anything, but not the pain of losing you. I want to make you the happiest women in the Universe … Priti Sinvi, will you marry me?"

"But what about children?"

"I'll post you my sperm."

"Oh, George, I do!" ∎

↻ NATURE.COM
Discuss this story
online at:
go.nature.com/emipyy

Anand Odhav Naranbhai *is a third-year medical student at the Nelson R. Mandela School of Medicine at the University of KwaZulu Natal in Durban, South Africa.*

JACEY

LETTER

Suppression of inflammation by a synthetic histone mimic

Edwige Nicodeme[1]*, Kate L. Jeffrey[2]*, Uwe Schaefer[2]*, Soren Beinke[3]*, Scott Dewell[4], Chun–wa Chung[5], Rohit Chandwani[2], Ivan Marazzi[2], Paul Wilson[5], Hervé Coste[1], Julia White[5], Jorge Kirilovsky[1], Charles M. Rice[6], Jose M. Lora[3], Rab K. Prinjha[3], Kevin Lee[3] & Alexander Tarakhovsky[2]

Interaction of pathogens with cells of the immune system results in activation of inflammatory gene expression. This response, although vital for immune defence, is frequently deleterious to the host due to the exaggerated production of inflammatory proteins. The scope of inflammatory responses reflects the activation state of signalling proteins upstream of inflammatory genes as well as signal-induced assembly of nuclear chromatin complexes that support mRNA expression[1–4]. Recognition of post-translationally modified histones by nuclear proteins that initiate mRNA transcription and support mRNA elongation is a critical step in the regulation of gene expression[5–10]. Here we present a novel pharmacological approach that targets inflammatory gene expression by interfering with the recognition of acetylated histones by the bromodomain and extra terminal domain (BET) family of proteins. We describe a synthetic compound (I-BET) that by 'mimicking' acetylated histones disrupts chromatin complexes responsible for the expression of key inflammatory genes in activated macrophages, and confers protection against lipopolysaccharide-induced endotoxic shock and bacteria-induced sepsis. Our findings suggest that synthetic compounds specifically targeting proteins that recognize post-translationally modified histones can serve as a new generation of immunomodulatory drugs.

BET proteins BRD2, BRD3 and BRD4 (hereafter defined as BET) govern the assembly of histone acetylation-dependent chromatin complexes that regulate inflammatory gene expression[5–8]. This function of BET suggests the possibility of intervention with inflammatory gene expression by disrupting chromatin complexes essential for mRNA transcription, elongation and splicing.

The diversity of binding surfaces, created by differences in sequences surrounding the bromodomain acetyl-binding pocket of BET, and other bromodomain-containing proteins, provided a foundation for selective pharmacological targeting of BET[9,11–14]. Using an approach that uses the ability of synthetic compounds to bind selectively to individual proteins in cell lysates (see Supplementary Material), we identified compounds that interact with BET. One of these compounds, GSK525762A (Fig. 1a), henceforth referred to as I-BET, showed the highest affinity interaction with BET (Fig. 1). The crystal structure of I-BET bound to BRD4-bromodomain 1 (BD1) showed I-BET positioned at the acetyl-lysine (AcK)-binding pocket (Fig. 1b and Supplementary Fig. 1a, b). Hydrogen bonding interactions essential for binding of AcK to asparagine 140 and tyrosine 97 within the bromodomain was mimicked by the triazoyl ring of I–BET (Fig. 1b). The selectivity of I-BET interaction with BET was determined by the ZA hydrophobic channel and WPF shelf outside of the AcK binding pocket, where a conserved isoleucine or valine impose spatial constraints on the size of molecules that can gain access to the WPF shelf (Fig. 1b and Supplementary Fig. 1b, c). Indeed, an enantiomer

compound of I-BET (GSK525768A) had no activity towards BET (Fig. 1c, far right panel). The structural features of I-BET allow two molecules of I-BET to bind to the tandem bromodomains of BET with high affinity (dissociation constant K_d of 50.5–61.3 nM; Fig. 1c and Supplementary Fig. 1d, e). Moreover, I-BET could successfully compete with AcK within the recognition pocket of BET. Fluorescence resonance energy transfer (FRET) analysis demonstrated that I-BET displaced, with high efficacy (half-maximum inhibitory concentration IC_{50} of 32.5–42.5 nM), a tetra-acetylated H4 peptide that had been pre-bound to tandem bromodomains of BET (Fig. 1d and Supplementary Fig. 1e). I-BET is highly selective as it did not interact with other bromodomain-containing proteins from each arm of the phylogeny tree (Supplementary Fig. 1f) and had no activity towards a panel of 38 unrelated proteins (Supplementary Table 1).

Stimulation of bone marrow-derived macrophages (BMDMs) with lipopolysaccharide (LPS) upregulated numerous inflammatory genes (Fig. 2a). Pre-treatment of BMDMs with I-BET shortly before LPS stimulation resulted in the downregulation of 38 and 151 of the LPS-inducible genes at 1 and 4 h, respectively (Fig. 2a, b and Supplementary Table 2). I-BET suppressed the expression of key LPS-inducible cytokines and chemokines, including $Il6$, $Ifnb1$, $Il1b$, $Il12a$, $Cxcl9$ and $Ccl12$. The inhibitory effect of I-BET on the expression of the IL-1β processing enzyme $Mefv$[15] underscored the potential of I-BET to control the IL-1β inflammatory circuit. Furthermore, diminished expression of transcription factors Rel, $Irf4$ and $Irf8$ point to the ability of I-BET to curtail the initial wave of inflammatory gene expression (Fig. 2b and Supplementary Table 2). In the absence of LPS stimulation, treatment of BMDMs with I-BET had a marginal effect on gene transcription (Supplementary Fig. 3 and Supplementary Table 3) and did not have an impact on the expression of $Tlr1–13$, $Myd88$, $Ticam1$, $Cd14$, $Mapk$, $Mapk2k$, $Map3k$ and $Map4k$ family members, $Ikbkb$, $Ikbke$ and $Ikbkg$ and $Aoah$[16], in unstimulated or LPS-treated macrophages at 1 h, that control LPS sensing and signalling (Supplementary Fig. 4). Furthermore, an unaltered pattern of LPS-induced ERK phosphorylation and IκBα degradation in I-BET-treated cells excluded the impact of I-BET on gene expression through dysregulation of Toll-like receptor 4 (TLR4)-dependent signalling (Supplementary Fig. 5). I-BET also had no effect on the expression of housekeeping genes or the viability of BMDMs (Supplementary Figs 4 and 5). The impact of I-BET on LPS-inducible gene expression is highly selective. The cytokine Tnf as well as chemokines $Ccl2–5$, $Cxcl1/2$ were not affected by I-BET (Supplementary Table 2 and Supplementary Fig. 6). This specificity and anti-inflammatory potential of I-BET has been validated by the similarity between the effects of I-BET treatment and siRNA-mediated BET knockdown on inflammatory gene expression (Supplementary Fig. 7). Notably, knockdown of BET suppressed the expression of Tnf that was resistant to I-BET (Supplementary Figs 6 and 7). This

result points to the existence of BET-recruiting mechanisms that are independent of BET interaction with acetylated histones. The existence of such a mechanism is supported by findings that show recruitment of BET to acetylated Rela or mediator complex[7,8,17,18]. Certain genes were upregulated by I-BET treatment but none of these have a well-established role in inflammation (Supplementary Table 2). The upregulation of *Brd2* and histone-encoding genes (Supplementary Table 2) may reflect the existence of a positive feedback mechanism where suppression of BET leads to a compensatory increase in the expression of chromatin proteins. The activating effect of I-BET on gene expression may also reflect the ability of BET to function not only as transcriptional co-activators but also as co-repressors[19].

The genome-wide analysis of the epigenetic states of LPS-inducible genes that were significantly suppressed or not affected by I-BET (sI-BET and naI-BET genes, respectively) provided a clue for the selective effect of I-BET on gene expression. Elevated basal levels of histone H3 and H4 acetylation (H3ac and H4ac) at the naI-BET gene promoters

indicated that naI-BET genes were already primed or actively involved in transcription (Fig. 3 and Supplementary Fig. 8). Indeed, the naI-BET gene promoters were associated with higher basal levels of H3K4me3 and RNA polymerase (Pol) II, including the elongation competent RNA Pol II, phosphorylated at serine 2 (RNA Pol II S2; Fig. 3). The important role of the primed/active state in defining the sensitivity to I-BET was underscored by the lack of I-BET effect on expression of housekeeping genes such as *Gapdh*, *Tubb5* and *Hprt* (Supplementary Fig. 4) that are characterized by high levels of H3ac, H4ac, H3K4me3 and RNA Pol II at their promoters[5]. Furthermore, an increase in overall histone acetylation levels caused by BMDM treatment with the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) was able to 'convert' sI-BET into naI-BET genes (Supplementary Fig. 9).

The primed and/or active transcription state of naI-BET genes before LPS stimulation was accomplished without recruitment of significant amounts of BET, thus reducing the likelihood of suppression of these genes by I-BET (Fig. 3). Furthermore, treatment with I-BET has less impact on BET association with naI-BET compared to sI-BET gene promoters in LPS-treated cells (Fig. 3). The mechanism of this
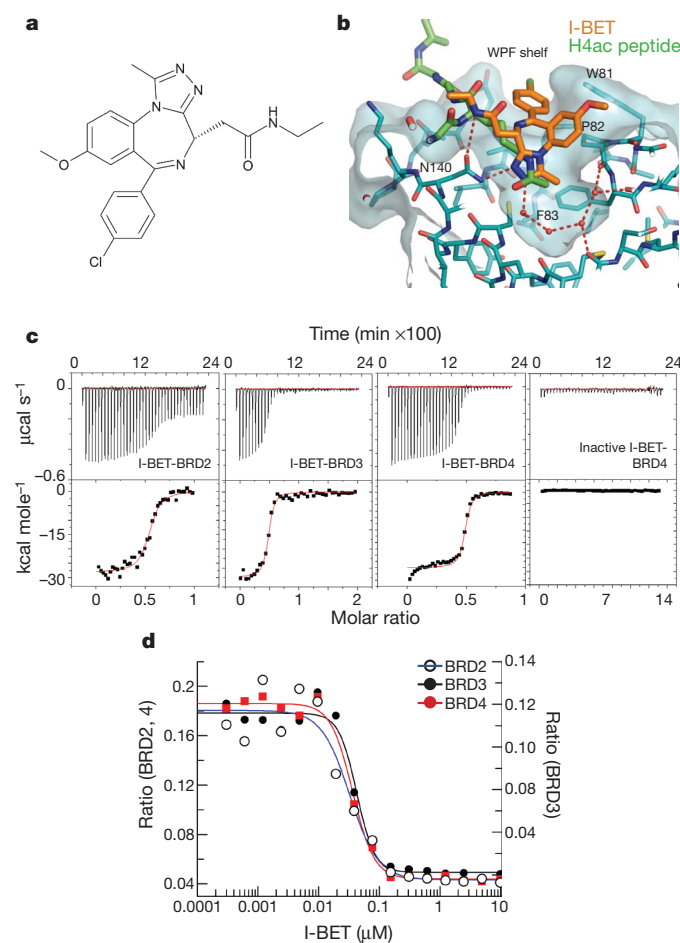
Figure 1 | I-BET is a selective antagonist of BET proteins. a, Chemical structure of GSK525762A (I-BET). **b**, Structure of I-BET (orange) bound to the acetyl-binding pocket of BRD4-BD1 overlaid with acetylated histone H4 peptide (H4ac, green). The 'WPF shelf' (W81, P82, F83) as well as the asparagine N140 essential for acetylated lysine (Kac) binding are indicated. **c**, I-BET binds with high affinity to BET proteins as determined by isothermal titration calorimetry (ITC) of tandem bromodomain fragments of BRD2 (1–473), BRD3 (1–434), BRD4 (1–477) interaction with I-BET or BRD4 (1–477) interaction with an inactive enantiomer of I-BET (inactive I-BET). Time courses of raw injection heats (upper panel) and normalized binding enthalpies, calculated using a single site binding model (Origin software, Microcal, lower panel), are shown. **d**, I-BET competes with H4ac peptide for bromodomain binding. Displacement of tetra-acetylated histone H4 peptide from bromodomains of BRD2 (blue), BRD3 (black) and BRD4 (red) by I-BET was determined by FRET analysis.
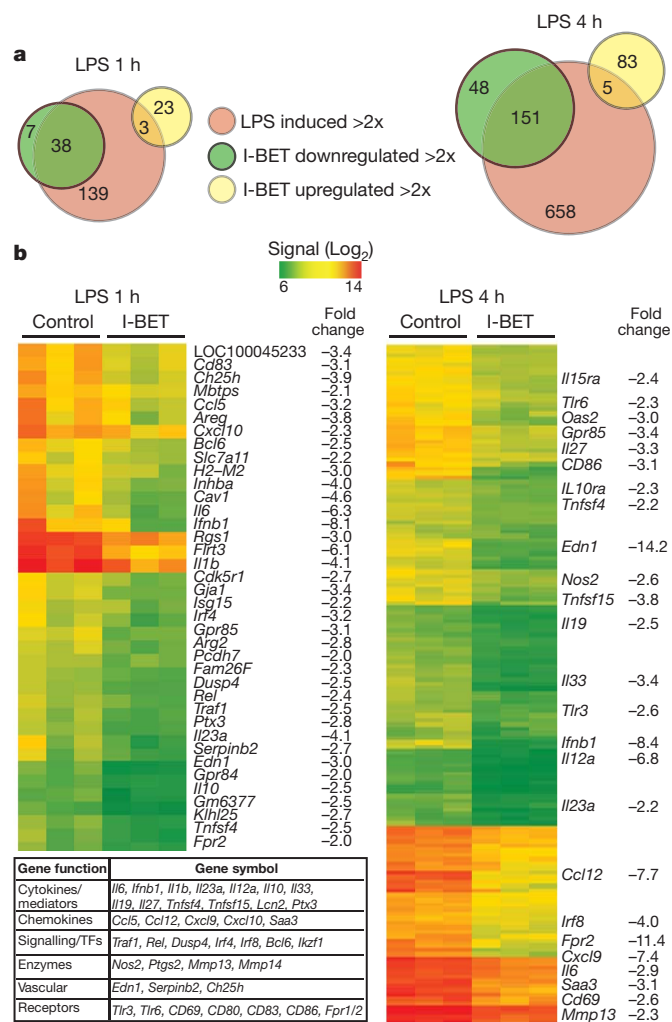
Figure 2 | I-BET suppresses a specific subset of LPS-inducible genes. a, Venn diagrams display the number of LPS-inducible (>twofold, red circles) genes that were suppressed (>twofold, green circles) or upregulated (>twofold, yellow circles) by I-BET (1 μM) treatment at 1 or 4 h after LPS stimulation (left and right panels) of BMDMs. **b**, Heat map representation of expression levels of genes that were downregulated by I-BET at 1 h (left panel) and 4 h (right panel) after LPS stimulation of three independent macrophage cultures. Scale ranges from a signal value of 2^6 (64, green) to 2^{14} (16,384, red). Fold-change values are listed. Table shows the distribution of downregulated genes into functional categories.
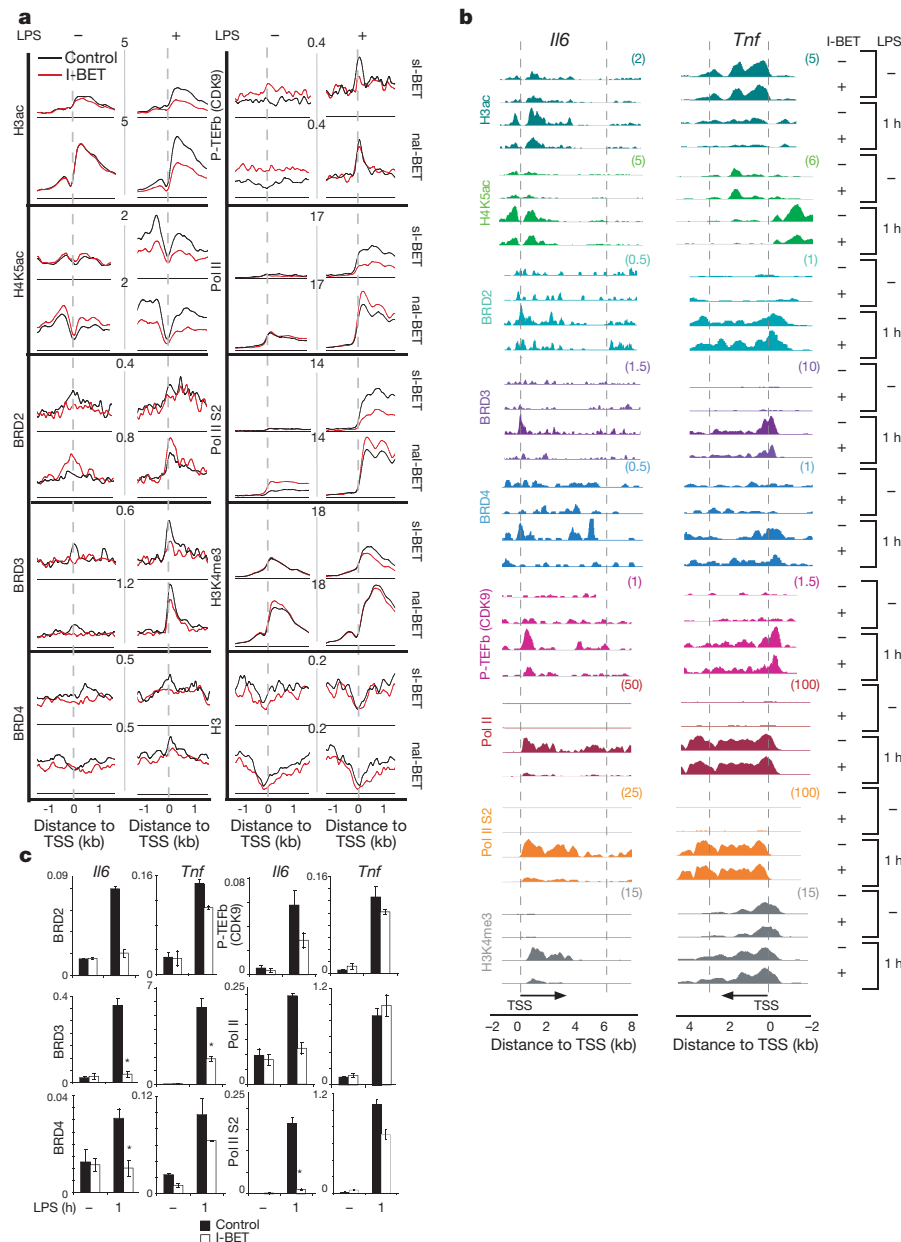
Figure 3 | Epigenetic profiles of genes suppressed or unaffected by I-BET in LPS-stimulated macrophages. a, Genome-wide epigenetic profiles of sI-BET or naI-BET genes in unstimulated or LPS-stimulated (1 h) macrophages pretreated with 5 μM of I-BET or a DMSO control. Analysed epigenetic marks are indicated. y-Axes represent the number of reads per million mapped reads. **b,** Epigenetic profiles of *Il6* and *Tnf*. The y-axes represent the average number of tags per gene per 25 base pairs per 1,000,000 mapped reads. Scale values are indicated in parentheses. **c,** The abundance of epigenetic marks on *Il6* and *Tnf* gene promoters was quantified by ChIP qPCR from four (BRD3, BRD4, Pol II and Pol II S2) or two (BRD2 and P-TEFb) independent experiments performed in triplicate. Error bars are s.e.m. of independent experiments or s.d. of representative experiments, respectively. Asterisks indicate $P < 0.05$ as determined by an unpaired t-test.

phenomenon may reflect higher LPS-induced H3ac and H4ac levels at naI-BET compared to sI-BET gene promoters before and after I-BET treatment (Fig. 3 and Supplementary Fig. 8). Additionally, some of the naI-BET genes may recruit BET via histone acetylation-independent mechanisms[17].

Treatment of BMDMs with I-BET affected not only the promoter-bound BET but also the levels of H3ac, H4K5ac, H4K8ac, H4K12ac and total H4ac on LPS-induced gene promoters (Fig. 3 and Supplementary Fig. 8). The mechanism of the negative impact of I-BET on histone H3 and H4 acetylation might be twofold. First, by preventing BET from binding to H3ac/H4ac, I-BET increases the accessibility of exposed H3ac/H4ac to HDACs. This model is supported by a nonenzymatic role of BRD4 in H4ac preservation in embryonic stem (ES) cells[20]. It is also possible that I-BET binding to BET prevents the

formation of multi-molecular complexes that contain histone acetyltransferases (HATs), other histone modifying enzymes, including lysine H3K4me3 methyltransferases, as well as the positive transcriptional elongation factor b (P-TEFb) and RNA Pol II[7,8,18,19]. This model is consistent with diminished levels of P-TEFb, that contributes to mRNA elongation by RNA Pol II phosphorylation[5,21,22], and reduced amounts of H3K4me3 and RNA Pol II at sI-BET genes (Fig. 3). The possible direct impact of I-BET on H3ac/H4ac through inhibition of bromodomain-containing HATs was excluded by the inability of I-BET to suppress the activity of the most common HATs such as pCAF, p300, GCN5 and CBP (also known as Kat2b, Ep300, Kat2a and Crebbp, respectively; Supplementary Fig. 10).

The features of sI-BET and naI-BET genes assessed by the genome-wide analysis were mirrored by the epigenetic states of selected sI-BET
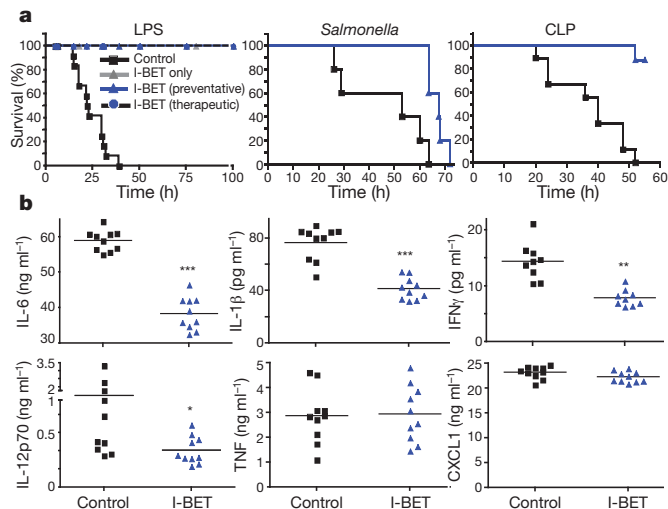
Figure 4 | I-BET suppresses inflammation *in vivo*. a, Kaplan–Meier survival curves of: LPS-treated C57BL/6 mice (5 mg per kg, i.p., *n* = 12 per group) that were injected i.v. with a solvent control (black squares) or 30 mg per kg of I-BET 1 h before (blue triangles) or 1.5 h after (blue circles) LPS administration (left panel); mice injected i.v. with heat-killed *Salmonella typhimurium*, strain IR71 (5×10^9 per kg, *n* = 10 per group) (middle panel); or mice subjected to caecal ligation puncture (CLP) procedure that were administered a solvent control or 30 mg per kg of I-BET twice a day for 2 days (*n* = 8 per group) (right panel). **b**, Serum titres of indicated cytokines were measured by ELISA (*n* = 10 per group). Mice received a solvent control (black squares) or I-BET (blue triangles) 1 h before LPS injection and samples were collected at 2 h after LPS treatment. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$ as determined by unpaired *t*-test.

and naI-BET genes. Following I-BET treatment, the promoter of the siBET gene *Il6* showed a marked reduction in BET recruitment and diminished levels of associated H3K4me3, P-TEFb, RNA Pol II and RNA Pol II S2 (Fig. 3b, c). In contrast to *Il6*, the naI-BET gene *Tnf* showed higher accumulation of BRD2, BRD3 and to a lesser extent BRD4, around the transcriptional start site (TSS) following I-BET treatment (Fig. 3b, c). The relatively higher BET levels at the *Tnf* locus were associated with largely unaffected levels of P-TEFb, RNA Pol II and RNA Pol II S2 (Fig. 3b, c). In support of distinct epigenetic states between sI-BET and naI-BET genes, the sI-BET gene *Il1b* had reduced BET accumulation at its TSS that resulted in a drop of P-TEFb, Pol II and Pol II S2 levels. In contrast, the epigenetic landscape of the naI-BET gene *Nfkbia* displayed little change in response to I-BET (Supplementary Fig. 11).

The selectivity of gene responses to I-BET correlated inversely with the timing of LPS-induced gene activation. Opposite to early stimulated (primary response) naI-BET genes, the majority of sI-BET genes, with the exception of *Il1b*, belong to the category of secondary response genes (SRG) that become upregulated at later points of macrophage activation (Supplementary Fig. 12a, c). Most of the sI-BET genes, as well as *Il1b*, were characterized by low basal levels of H3ac/H4ac, H3K4me3, RNA Pol II, as well as low CpG content of their promoters (Fig. 3, Supplementary Figs 8 and 12b). The latter feature conveys higher stability to promoter-associated nucleosomes that generates a selective barrier for transcriptional activation of secondary response genes[23,24]. It is likely that suppression of BET recruitment as well as reduction in H3ac/H4ac and H3K4me3 by I-BET aggravates the already non-permissive transcriptional state of the sI-BET genes and reduces the probability of their expression, thus defining the selectivity of I-BET.

The suppression of key inflammatory genes by I-BET suggested a potent ability of the compound to treat inflammatory conditions *in vivo*. The serum titres of intravenously (i.v.) administered I-BET remain within the effective concentrations for several hours after injection (Supplementary Fig. 13). Injection of I-BET in mice before the initiation of LPS- or heat-killed *Salmonella typhimurium*-induced

endotoxic shock was able to prevent or attenuate death of mice (Fig. 4a left and middle panel). Most promisingly for therapeutic applications, a single dose of I-BET applied at 1.5 h after LPS injection, at the time when mice started to develop symptoms of inflammatory disease, cured the mice (Fig. 4a, left panel). Furthermore, in mice that suffer from polymicrobial peritonitis and sepsis caused by caecal ligation and puncture (CLP), twice-daily injections of I-BET for 2 days protected mice against death caused by sepsis (Fig. 4a, right panel).

The marked therapeutic effect of I-BET on endotoxic shock and sepsis occurred despite unaltered serum TNF levels (Fig. 4b). As TNF is an established mediator of sepsis-associated inflammatory processes, the protective effect of I-BET on sepsis suggests the ability of I-BET to interfere not only with the expression of inflammatory proteins (Fig. 4b), but also with TNF-inducible gene expression. Indeed, treatment of BMDMs with I-BET suppressed TNF-inducible key pro-inflammatory cytokine (*Il1b*, *Il1a*) and chemokine genes (*Ccl5*, *Cxcl10*, *Cxcl2/3*) as well as vasoactive and lipid-related genes (*Pdgfb*, *Adora2b*, *Fabp3*) that contribute to sepsis pathogenesis (Supplementary Fig. 14a, b). Notably, similar to the sI-BET genes in LPS-treated BMDMs, the majority of sI-BET genes in TNF-treated cells fit into the secondary response gene category as assessed by epigenetic modifications and CpG content (Supplementary Fig. 14c).

In summary, we show the anti-inflammatory potential of the synthetic compound I-BET that, by interfering with binding of bromodomain-containing BET proteins to acetylated histones, disrupts the formation of the chromatin complexes essential for expression of inflammatory genes. The genes susceptible to I-BET share a common pattern of chromatin modifications at their promoters as well as low promoter CpG content. Suppression of inflammation by I-BET demonstrates the potential of drugs that interfere with protein binding to post-translationally modified histones to achieve a high level of selectivity and potency by exploiting the inherited epigenetic states of genes that contribute to specific physiological and pathological processes.

METHODS SUMMARY

I-BET is an optimized derivative of benzodiazepine compounds that were identified by high-throughput screening of activators of ApoA1-luciferase reporter in HepG2 cells as described in Supplementary Information. The chemical synthesis of I-BET is described in Supplementary Information. The 1.6 Å crystal structure of BRD4-BD1 with I-BET was produced by soaking apo crystals in 2 mM I-BET for 4 days. Molecular replacement using 2oss.pdb gave excellent difference density at the acetylated binding site that allowed the ligand binding to be unambiguously modelled. Methods and statistics for data collection and refined coordinates are provided in Supplementary Information and deposited in the RCSB Protein Data Bank with PDB ID code 3P5O. Bone marrow-derived macrophages (BMDMs) were differentiated from a bone marrow cell suspension obtained from C57BL/6 mice as described in supplementary information. For microarray, qPCR (quantitative PCR) and ChIP (chromatin immunoprecipitation) analyses, BMDMs were pre-incubated with 1 μM or 5 μM of I-BET, DMSO or an inactive I-BET compound for 30 min before LPS (100 ng ml^{-1}) or TNF (50 ng ml^{-1}) stimulation. Microarray experiments were performed using Illumina MouseRef-8 v2.0 expression BeadChip kits (GEO accession code GSE21764). qPCR was performed using SYBR Green (Roche Lightcycler 480). ChIP was performed as described[25] and detailed in supplementary information. ChIP sequencing libraries were generated as described[26] (GEO accession code GSE21910). For LPS-induced endotoxic shock, 5 mg per kg of LPS was injected intraperitoneally (i.p.) into age-matched C57BL/6 mice. Heat-killed *Salmonella typhimurium* (IR715; 5×10^9 per kg) was injected intravenously. Caecal ligation puncture (CLP) was performed as described[27]. For *in vivo* experiments I-BET or a solvent control (20% beta-cyclodextrin, 2% DMSO in 0.9% saline) were given via retro-orbital or tail vein injection (CLP) at a dose of 30 mg per kg.

1. Medzhitov, R. & Horng, T. Transcriptional control of the inflammatory response. *Nature Rev. Immunol.* **9,** 692–703 (2009).
2. Smale, S. T. Selective transcription in response to an inflammatory stimulus. *Cell* **140,** 833–844 (2010).

3. Natoli, G. Control of NF-κB-dependent transcriptional responses by chromatin organization. *Cold Spring Harb. Perspect. Biol.* **1,** a000224 (2009).
4. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416,** 499–506 (2002).
5. Hargreaves, D. C., Horng, T. & Medzhitov, R. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138,** 129–145 (2009).
6. LeRoy, G., Rickards, B. & Flint, S. J. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol. Cell* **30,** 51–60 (2008).
7. Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19,** 523–534 (2005).
8. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19,** 535–545 (2005).
9. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14,** 1025–1040 (2007).
10. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
11. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Rev. Mol. Cell Biol.* **8,** 983–994 (2007).
12. Huang, H. *et al.* Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct. Biol.* **7,** 57 (2007).
13. Liu, Y. *et al.* Structural basis and binding properties of the second bromodomain of Brd4 with acetylated histone tails. *Biochemistry* **47,** 6403–6417 (2008).
14. Vollmuth, F., Blankenfeldt, W. & Geyer, M. Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution. *J. Biol. Chem.* **284,** 36547–36556 (2009).
15. Gavrilin, M. A. *et al.* Pyrin critical to macrophage IL-1β response to *Francisella* challenge. *J. Immunol.* **182,** 7982–7989 (2009).
16. Hagen, F. S. *et al.* Expression and characterization of recombinant human acyloxyacyl hydrolase, a leukocyte enzyme that deacylates bacterial lipopolysaccharides. *Biochemistry* **30,** 8415–8423 (1991).
17. Huang, B., Yang, X. D., Zhou, M. M., Ozato, K. & Chen, L. F. Brd4 coactivates transcriptional activation of NF-κB via specific binding to acetylated RelA. *Mol. Cell. Biol.* **29,** 1375–1387 (2009).
18. Jiang, Y. W. *et al.* Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl Acad. Sci. USA* **95,** 8538–8543 (1998).
19. Denis, G. V. *et al.* Identification of transcription complexes that contain the double bromodomain protein Brd2 and chromatin remodeling machines. *J. Proteome Res.* **5,** 502–511 (2006).
20. Nishiyama, A., Dey, A., Miyazaki, J. & Ozato, K. Brd4 is required for recovery from antimicrotubule drug-induced mitotic arrest: preservation of acetylated chromatin. *Mol. Biol. Cell* **17,** 814–823 (2006).
21. Marshall, N. F., Peng, J., Xie, Z. & Price, D. H. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J. Biol. Chem.* **271,** 27176–27183 (1996).
22. Sims, R. J. III, Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18,** 2437–2468 (2004).
23. Ramirez-Carrozzi, V. R. *et al.* Selective and antagonistic functions of SWI/SNF and Mi-2β nucleosome remodeling complexes during an inflammatory response. *Genes Dev.* **20,** 282–296 (2006).
24. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138,** 114–128 (2009).
25. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature Protocols* **1,** 729–748 (2006).
26. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140,** 678–691 (2010).
27. Rittirsch, D., Huber-Lang, M. S., Flierl, M. A. & Ward, P. A. Immunodesign of experimental sepsis by cecal ligation and puncture. *Nature Protocols* **4,** 31–36 (2008).

Oxidant stress evoked by pacemaking in dopaminergic neurons is attenuated by DJ-1

Jaime N. Guzman[1], Javier Sanchez-Padilla[1], David Wokosin[1], Jyothisri Kondapalli[2], Ema Ilijic[1], Paul T. Schumacker[2] & D. James Surmeier[1]

Parkinson's disease is a pervasive, ageing-related neurodegenerative disease the cardinal motor symptoms of which reflect the loss of a small group of neurons, the dopaminergic neurons in the substantia nigra pars compacta[1] (SNc). Mitochondrial oxidant stress is widely viewed as being responsible for this loss[2], but why these particular neurons should be stressed is a mystery. Here we show, using transgenic mice that expressed a redox-sensitive variant of green fluorescent protein targeted to the mitochondrial matrix, that the engagement of plasma membrane L-type calcium channels during normal autonomous pacemaking created an oxidant stress that was specific to vulnerable SNc dopaminergic neurons. The oxidant stress engaged defences that induced transient, mild mitochondrial depolarization or uncoupling. The mild uncoupling was not affected by deletion of cyclophilin D, which is a component of the permeability transition pore, but was attenuated by genipin and purine nucleotides, which are antagonists of cloned uncoupling proteins. Knocking out *DJ-1* (also known as *PARK7* in humans and *Park7* in mice), which is a gene associated with an early-onset form of Parkinson's disease, downregulated the expression of two uncoupling proteins (UCP4 (SLC25A27) and UCP5 (SLC25A14)), compromised calcium-induced uncoupling and increased oxidation of matrix proteins specifically in SNc dopaminergic neurons. Because drugs approved for human use can antagonize calcium entry through L-type channels, these results point to a novel neuroprotective strategy for both idiopathic and familial forms of Parkinson's disease.

Calcium entry through L-type channels in SNc dopaminergic neurons occurs throughout the pacemaking cycle[3,4], contrasting them with neighbouring dopaminergic neurons in the ventral tegmental area (VTA), which are much less affected in Parkinson's disease[5] (Fig. 1a, b). Although prominent, this influx is not necessary for pacemaking, because treatment with the dihydropyridine L-type channel antagonist isradipine eliminates cytosolic calcium oscillations but leaves pacemaking intact[6] (Fig. 1a).

Calcium entry during pacemaking comes at a metabolic cost, as it must be extruded by ATP-dependent processes. This demand is met primarily by mitochondria through oxidative phosphorylation. Superoxide and reactive oxygen species are by-products of oxidative phosphorylation, raising the possibility that calcium entry creates mitochondrial oxidant stress. To determine whether this was the case, we generated[7] transgenic mice expressing a redox-sensitive variant of green fluorescent protein (roGFP) with a mitochondrial-matrix-targeting sequence (mito-roGFP). To limit expression to monoaminergic neurons, we expressed mito-roGFP under the control of the tyrosine hydroxylase promoter (TH-mito-roGFP; Fig. 1c). Dopaminergic neurons in the SNc and the adjacent VTA from these mice robustly expressed mito-roGFP that co-localized with mitochondrial markers (Fig. 1c–e and Supplementary Fig. 1), providing a reversible, quantitative means of monitoring the oxidation of mitochondrial matrix proteins. Because the expression of mito-roGFP was restricted to a small set of neurons, it was possible to monitor the mitochondrial redox

state in individual neurons deep in brain slices from young adult mice using two-photon laser scanning microscopy (2PLSM).

In VTA dopaminergic neurons, the basal oxidation of mito-roGFP was very low (Fig. 1f). In contrast, the oxidation of mito-roGFP was significantly higher in neighbouring SNc dopaminergic neurons (Fig. 1g, h). In juvenile SNc dopaminergic neurons, where pacemaking is similar to that of VTA neurons, mitochondrial oxidant stress also was low (Fig. 1h). To verify that this stress was not an artefact of brain slicing, we generated a transgenic mouse that expressed mito-roGFP under the control of the cytomegalovirus promoter, yielding robust neuronal expression in the cerebral cortex, striatum and hippocampus. In brain slices from these mice, principal neurons in each of these regions were devoid of any significant mitochondrial oxidation (Supplementary Fig. 2), demonstrating that slicing per se did not create oxidant stress.

What did contribute significantly to the mitochondrial oxidant stress was calcium influx through plasma membrane L-type channels. Antagonizing L-type channels drastically lowered the extent of mito-roGFP oxidation (Fig. 1g, h), as did slowing pacemaking by cooling (Fig. 1i). L-type channel antagonists had no effect on the oxidation of matrix proteins in neighbouring VTA dopaminergic neurons (Fig. 1h). Blocking calcium entry into mitochondria from the cytoplasm with Ru360[8] diminished roGFP oxidation (without affecting pacemaking) (Fig. 1h), suggesting that it helped to drive oxidative phosphorylation[9].

Loss-of-function mutations in DJ-1 are linked to an autosomal recessive, early-onset form of Parkinson's disease[10]. Although DJ-1 is not an antioxidant enzyme itself, it is redox sensitive and participates in signalling cascades made active by mitochondrial superoxide generation. To examine its role in SNc dopaminergic neurons, *DJ-1* knockout mice were crossed with the TH-mito-roGFP mice. SNc dopaminergic neurons from these mice had normal pacemaking and oscillations in intracellular calcium concentration (Fig. 2a). However, basal mito-roGFP oxidation was nearly complete at physiological temperatures in these neurons, so we re-examined cells at a lower temperature. These studies confirmed the robust difference in oxidation between wild-type and *DJ-1* knockout neurons seen at higher temperatures (Fig. 2b, c). This difference was virtually abolished by antagonism of L-type calcium channels (Fig. 2b, c). In contrast, the mitochondria in neighbouring VTA dopaminergic neurons were unaffected by *DJ-1* deletion (Fig. 2d).

A clue about the role of DJ-1 in attenuating mitochondrial oxidant stress came from measurements of the inner mitochondrial membrane (IMM) potential with the cationic dye tetramethylrhodamine methylester (TMRM) (Fig. 3a and Supplementary Movie 1). In VTA dopaminergic neurons, TMRM fluorescence was robust and stable for long periods (Fig. 3b). In contrast, mitochondrial TMRM fluorescence in neighbouring SNc dopaminergic neurons repeatedly fell and then rose back to peak values, indicating that mitochondria were transiently depolarizing (Fig. 3b and Supplementary Movie 1). This 'flickering' was stable for long periods (>60 min) and was peculiar to SNc dopaminergic neurons, suggesting that it was not a product of the preparation

[1]Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. [2]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.
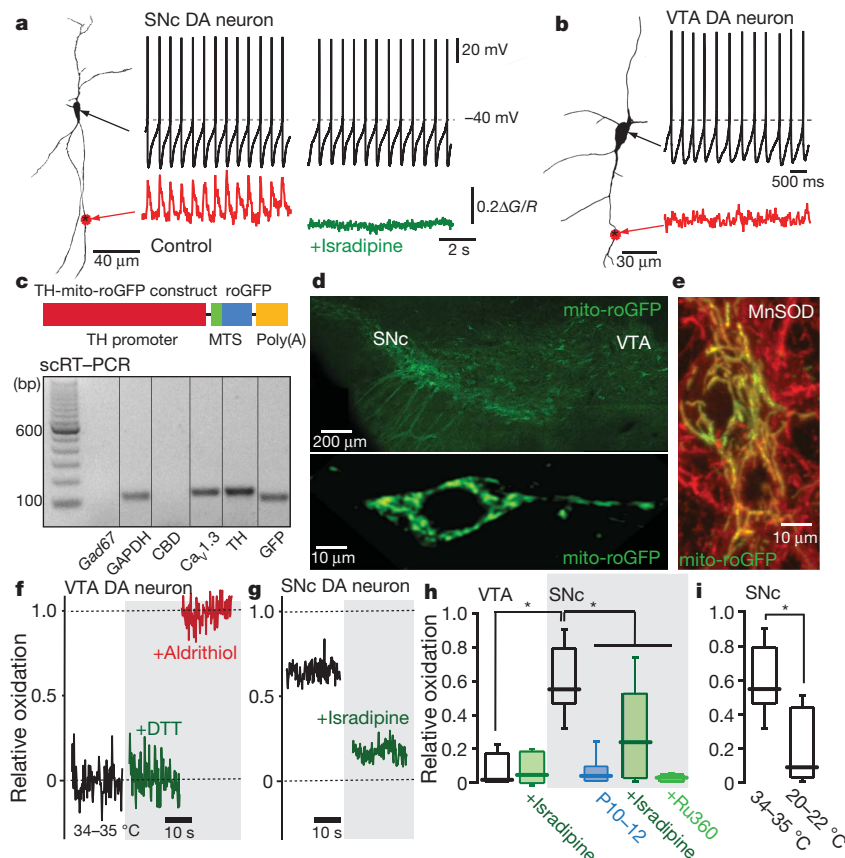
Figure 1 | Calcium influx through L-type calcium channels during pacemaking increases mitochondrial oxidant stress in SNc dopaminergic neurons. a, Somatic whole-cell recording from a SNc dopaminergic (DA) neuron (shown to the left as a projection image) in pacemaking mode. At the bottom of the panel, a 2PLSM measurement of dendritic Fluo-4 fluorescence (red trace) is shown. To the right are shown a similar set of measurements after application of isradipine; we note the absence of a change in pacemaking rate but the loss of dendritic calcium transients (green trace) ($P < 0.05$, $n = 10$ neurons). **b**, A similar set of measurements in a VTA dopaminergic neuron; these neurons consistently lacked dendritic calcium oscillations ($n = 6$ neurons). **c**, Schematic of the TH-mito-roGFP construct. Below this is a single-cell reverse-transcription PCR (scRT–PCR) analysis of an mito-roGFP-expressing SNc dopaminergic neuron showing expression of tyrosine hydroxylase (TH) and $Ca_V1.3$ calcium channel mRNA but not calbindin (CBD) or *Gad67* (*Gad1*) mRNA; similar results were obtained in all five neurons examined. bp, base pairs. MTS, matrix targeting sequence. **d**, Top: a low-magnification image of the mesencephalon of a transgenic TH-mito-roGFP mouse showing expression in SNc and VTA neurons. Bottom: a higher-magnification image of a SNc neuron showing cytoplasmic but not nuclear labelling. **e**, Overlay of manganese superoxide dismutase (MnSOD) immunostaining (red) showing co-localization with mito-roGFP in cultured roGFP SNc neurons. **f**, Mito-roGFP measurements from a VTA neuron; before (control, black trace) and after application of dithiothreitol (DTT; green trace)

and Aldrithiol (red trace). Relative oxidation is plotted as described in Methods. **g**, Mito-roGFP measurements from a SNc dopaminergic neuron (black trace) revealed a higher basal oxidation; treatment with isradipine reduced mitochondrial oxidation (green trace). **h**, Left: box plots summarizing mean redox measurements in control ($n = 9$) and isradipine-treated ($n = 5$) VTA dopaminergic neurons, and in control ($n = 14$), isradipine-treated ($n = 9$) and Ru360-treated ($n = 8$) SNc dopaminergic neurons; SNc neurons were significantly more oxidized than VTA neurons ($P < 0.05$), and both isradipine and Ru360 significantly reduced oxidation in SNc neurons ($P < 0.05$). SNc dopaminergic neurons from young mito-roGFP mice (postnatal days (P) 10–12) were consistently less oxidized than adult neurons ($P < 0.05$, $n = 8$). In these box plots (and those in Figs 2–4), the median is plotted as the central bar of the box, the upper and lower quartiles are the edges of the box, and the whiskers show the extremes of the distribution except for outliers, which are indicated using an asterisk. **i**, Box plots summarizing mean mito-roGFP measurements in SNc dopaminergic neurons in brain slices held at near-physiological temperature (34–35 °C; $n = 14$) and at room temperature (20–22 °C; $n = 5$); there was significantly less oxidation at room temperature ($P < 0.05$). Statistical significance in all plots (with stated P values) is shown by an asterisk and was determined using a non-parametric test for comparing multiple groups (Kruskal–Wallis analysis of variance with Dunnet's *post hoc* test).

(Supplementary Fig. 2). Judging by a Nernst equation relating IMM potential to the ratio of mitochondrial to nuclear TMRM fluorescence[11], it seemed that the flickering in mitochondrial potential was modest, corresponding to a IMM depolarization of only 20–30 mV.

Antagonizing plasma membrane L-type calcium channels with isradipine drastically reduced the rate of flickering (Fig. 3c and Supplementary Movie 1), as did blocking calcium entry into the mitochondria with Ru360 (Fig. 3c). However, flickering was also attenuated by scavenging reactive oxygen species with the cell-permeable antioxidant *N*-(2-mercaptopropionyl)-glycine (Fig. 3d), suggesting that oxidant stress created by calcium entry, rather than calcium per se, was responsible for the mitochondrial flickering. This inference was consistent with the fact that mitochondrial flickering was roughly ten

times slower than the oscillation in cytosolic calcium concentration (compare with Figs 1a and 3b).

Superoxide generation is known to trigger the opening of two types of ion channel that depolarize the IMM. One of these is the permeability transition pore[12]. However, TMRM flickering was normal in SNc dopaminergic neurons lacking cyclophilin D (Fig. 3e). Because cyclophilin D is one of its key modulators, this argues against a role for the permeability transition pore. Cyclosporin A, which is known to antagonize the permeability transition pore, decreased IMM flickering but also slowed or stopped pacemaking (and calcium influx), making it an unreliable diagnostic tool.

Uncoupling proteins (UCPs) are another class of mitochondrial ion channel whose open probability is increased by superoxide[13]. UCP
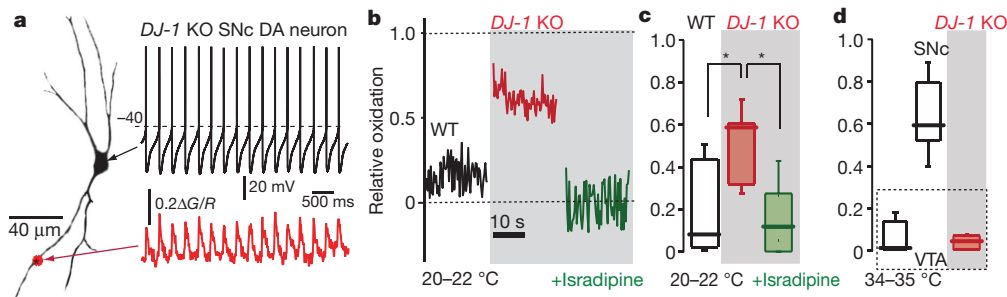
Figure 2 | Oxidant stress is elevated in SNc dopaminergic neurons from *DJ-1* knockout mice. a, Somatic whole-cell recording from a SNc dopaminergic neuron in a brain slice from a *DJ-1* knockout (KO) mouse, showing normal pacemaking (top) and intracellular calcium oscillations (bottom); similar results were obtained in all five neurons examined. **b**, Mitochondrial mito-roGFP oxidation in *DJ-1* knockout neurons was higher (red trace) than in control (wild-type, WT) SNc dopaminergic neurons (black trace); isradipine pre-treatment normalized oxidation of mito-roGFP (green trace). Experiments were done at 20–22 °C. **c**, Box plot summarizing mean mito-roGFP measurements in wild-type SNc neurons ($n = 9$), *DJ-1* knockout SNc neurons ($n = 6$) and *DJ-1* knockout neurons after isradipine pre-treatment ($n = 7$); differences between wild-type and *DJ-1* knockout neurons were significant ($P < 0.05$), as were differences between knockouts with and without isradipine treatment ($P < 0.05$). **d**, Box plot summarizing mean mito-roGFP measurements from wild-type VTA dopaminergic neurons ($n = 9$), wild-type SNc dopaminergic neurons ($n = 14$) and *DJ-1* knockout VTA dopaminergic neurons (red box; $n = 4$) at 34–35 °C. VTA dopaminergic neurons were unaffected by *DJ-1* deletion ($P > 0.05$).

opening modestly decreases the IMM potential[14], making it a plausible mediator of the drop in IMM potential inferred from the TMRM measurements. Five UCPs have been cloned, three of which are robustly expressed in the SNc[15] (UCP2, UCP4 and UCP5). Application of the UCP antagonist genipin[16] significantly decreased the frequency and amplitude of mitochondrial flickering (Fig. 3e), as did dialysing neurons with ATP (Supplementary Fig. 3), providing support for UCP mediation[17]. The modest depolarization brought about by UCP activation is thought to diminish superoxide generation without significantly compromising ATP production, creating a protective, negative-feedback system to complement enzymatic defences against reactive oxygen species[17,18]. If this were the case, blocking UCPs with genipin should cause superoxide concentrations and oxidation of mitochondrial proteins to increase. Indeed, genipin significantly elevated oxidation of mito-roGFP in SNc dopaminergic neurons, whereas it had no effect on mitochondrial oxidation in VTA dopaminergic neurons, where the UCP defence was not engaged (Fig. 4a).

In *DJ-1* null SNc dopaminergic neurons, mitochondrial flickering was reduced in amplitude and frequency, suggesting that their UCP defences were compromised (Fig. 4b). Because DJ-1 is largely found outside mitochondria, it is not likely to have a direct role in gating UCPs[19]. Another function of DJ-1 is to upregulate the expression of antioxidant proteins in response to stress[10,20,21]. Quantitative analysis of messenger RNA from the SNc of *DJ-1* knockout mice revealed normal concentrations of *Ucp2* mRNA (Fig. 4c), suggesting that UCP2 was not a participant in flickering—a conclusion consistent with examination of neurons from *Ucp2* knockout mice (data not shown). In contrast, *Ucp4* and *Ucp5* mRNAs were downregulated in the SNc of *DJ-1* knockout mice (Fig. 4c), bringing expression levels down to those found in the unstressed VTA. UCP expression levels in the cerebral cortex, striatum and hippocampus were largely unchanged by *DJ-1* deletion (Supplementary Fig. 4). The expression of mRNAs for the antioxidant enzymes manganese superoxide dismutase, glutathione peroxidase and catalase also were largely unchanged in the SNc of *DJ-1* knockout mice, but immunoreactivity for manganese superoxide dismutase *in situ* was higher (Supplementary Fig. 5). These data suggest that the elevation in mitochondrial oxidant stress in SNc dopaminergic neurons lacking DJ-1 was not due to lowered expression of antioxidant enzymes but rather to diminished UCP4 and UCP5 expression, blunting UCP-mediated mitochondrial uncoupling in response to calcium-induced stress.

Collectively, our studies demonstrate that in mature SNc dopaminergic neurons there is a basal mitochondrial oxidant stress. This oxidant stress is not a consequence of old age, pathology or the experimental preparation, but rather one of a neuronal design that engages L-type calcium channels during autonomous pacemaking. Mitochondrial oxidant stress of this sort has long been thought to be central to the aetiology of Parkinson's disease[2]. However, there has been no explanation for why this stress should be greater in a small population of mesencephalic neurons. Our results fill this gap (Fig. 4d). The basal oxidant stress in SNc dopaminergic neurons was amplified by deletion of *DJ-1*, a gene associated with an early-onset, familial form of Parkinson's disease. In agreement with its putative role in regulating oxidant defences[10], deletion of *DJ-1* decreased the abundance of *Ucp5* and *Ucp4* mRNA in the SNc and compromised mitochondrial uncoupling in response to oxidant stress. Although the hypothesis that the effects of *DJ-1* deletion on SNc dopaminergic neurons are mediated by reduced UCP expression remains to be definitively tested, this scenario provides an example of how a mutation in a widely expressed gene might manifest itself only in subpopulation of neurons. Our data also provide a potential explanation for the unusual accumulation of mitochondrial DNA mutations with age in SNc dopaminergic neurons[22,23]. These mutations, which are attributable to accumulated superoxide exposure, diminish mitochondrial competence and promote phenotypic decline, proteostatic impairments and death[24,25].

Because the mitochondrial oxidant stress in SNc dopaminergic neurons was attenuated by exposure to dihydropyridines, our results also point to a potential neuroprotective strategy in both idiopathic and familial Parkinson's disease. Dihydropyridines have a long history of safe use in humans and have good pharmacokinetic properties, including the ability to cross the blood–brain barrier[26]. Recent epidemiological studies support their potential value, showing that dihydropyridine use substantially reduces the risk of Parkinson's disease[27,28].

METHODS SUMMARY

Midbrain slices were obtained from mice between postnatal days 21 and 30. Mice were handled according to guidelines established by the Northwestern University Animal Care and Use Committee, the National Institutes of Health and the Society for Neuroscience. We visualized midbrain slices using an infrared differential interference contrast video microscopy system (for patch-clamp recording) and imaged them using 2PLSM to measure calcium transients, the mitochondrial membrane potential (using TMRM dye) or the mito-roGFP signal. SNc or VTA neurons were filled with Alexa594 and Fluo-4 and calcium transients were imaged as described previously[6]. We calculated the mitochondrial membrane potential using a Nernst equation describing the distribution of TMRM[11]. Transgenic mice were generated by conventional approaches with a roGFP2 construct containing the tyrosine hydroxylase promoter and a mitochondrial-matrix-targeting sequence. Relative oxidation of mito-roGFP was determined from fluorescence measurements after fully reducing mitochondria with dithiothreitol and then fully oxidizing with Aldrithiol. Because the calibrated signal becomes independent of the absolute expression level of mito-roGFP, this strategy allows cell-to-cell comparisons. Results in the main body of the paper were derived from a single line of
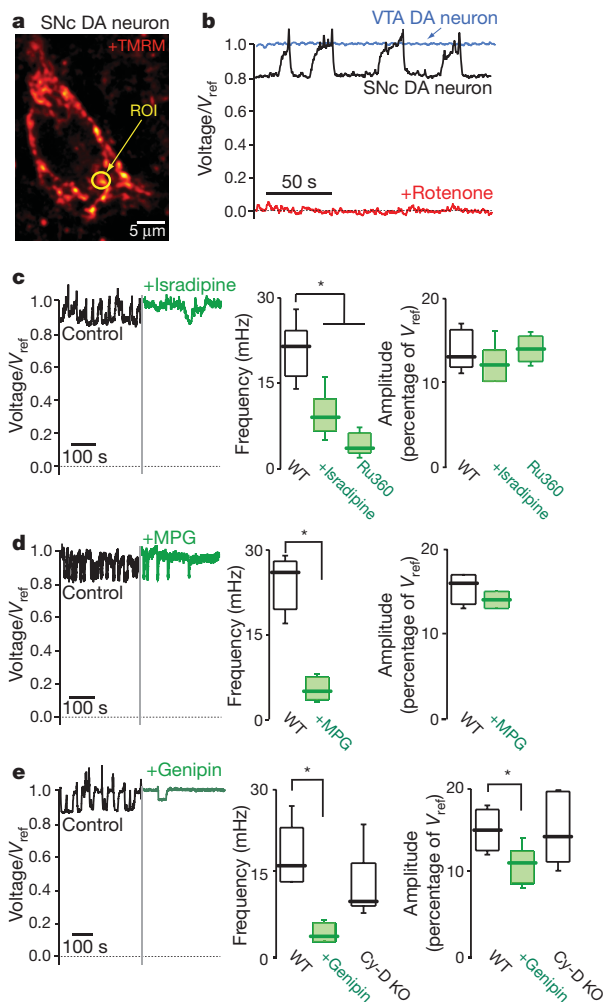
Figure 3 | Mitochondrial flickering is dependent on superoxide production and recruitment of mitochondrial uncoupling proteins. a, SNc dopaminergic neuron in a brain slice incubated with TMRM to label mitochondria; a region of interest (ROI) from which fluorescence measurements were taken is shown (yellow circle). **b**, Representative fluorescence time series from a SNc dopaminergic neuron before (black trace) and after (red trace) rotenone application; similar results were seen in all cells examined ($n > 20$). For comparison, a time series from a typical VTA dopaminergic neuron is shown (blue trace; $n = 10$). **c**, Left: TMRM fluorescence measurements before (black trace) and after (green trace) bath application of isradipine. Right: box plots summarizing the mean frequency of flickering in control cells before and after application of isradipine ($n = 5$) or Ru360 ($n = 6$), both of which significantly slowed flickering frequency ($P < 0.05$), and box plots summarizing the mean amplitude of the relative voltage change inferred from the fluorescence measurement; the amplitudes were similar in all conditions ($P > 0.05$). **d**, Left: fluorescence time series before and after bath application of N-(2-mercaptopropionyl)-glycine (MPG). Right: box plots summarizing mean frequency and amplitude measurements ($n = 5$); MPG significantly reduced the frequency ($P < 0.05$) but not the amplitude ($P > 0.05$) of flickering. **e**, Left: fluorescence time series before and after application of genipin. Right: box plots summarizing mean amplitude and frequency measurements ($n = 5$); genipin significantly decreased both parameters ($P < 0.05$). Both measurements were normal in cyclophilin D (Cy-D) knockouts ($P > 0.05$, $n = 10$).
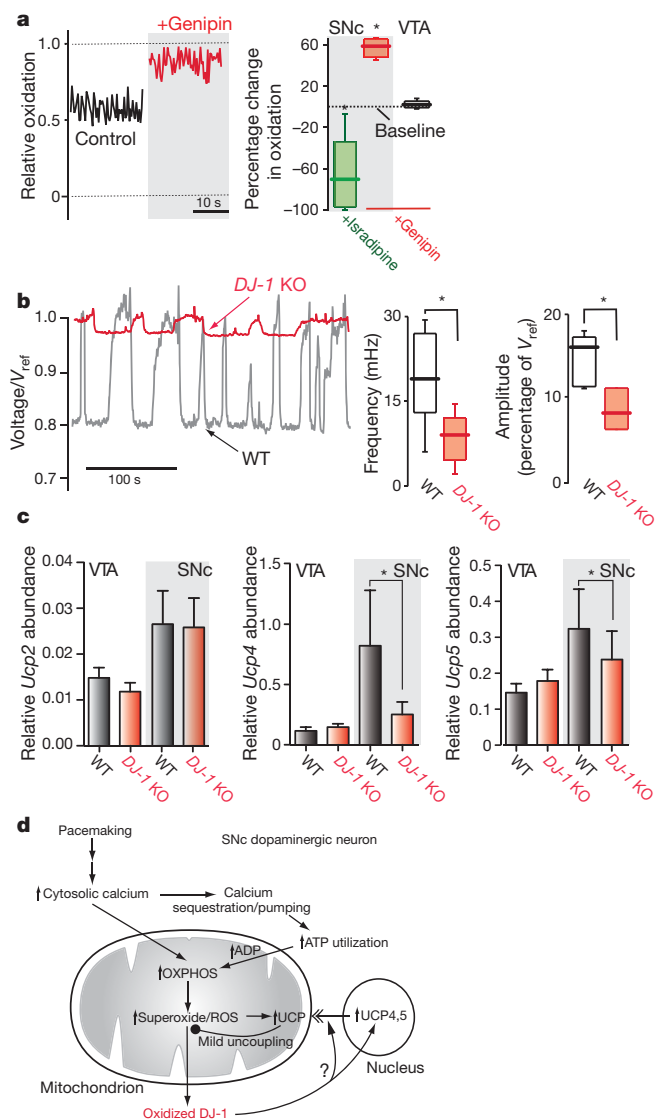
Figure 4 | Loss of DJ-1 attenuated UCP-dependent flickering in mitochondrial membrane potential. a, Left: mito-roGFP measurements in a SNc dopaminergic neuron (as in Fig. 1) before (black trace) and after (red trace) application of genipin. Right: box plots summarizing mean mito-roGFP measurements following application of isradipine (green box; $n = 9$) or genipin (red box; $n = 6$) to SNc dopaminergic neurons; isradipine significantly decreased oxidation, whereas genipin increased oxidation ($P < 0.05$). Also show is a box plot of mito-roGFP measurements from VTA dopaminergic neurons following genipin application; genipin had no effect on these measurements ($P > 0.05$, $n = 5$). **b**, Left: TMRM fluorescence measurement from a wild-type SNc dopaminergic neuron (black trace) and a *DJ-1* knockout neuron (red trace). Right: box plots of mean frequency and amplitude data from wild-type ($n = 21$) and *DJ-1* knockout neurons ($n = 7$); both amplitude and frequency of flickering were decreased in *DJ-1* knockouts ($P < 0.05$). **c**, Quantitative PCR analysis of UCP expression in *DJ-1* knockout mice: *Ucp2* mRNA abundance (relative to GAPDH) was not significantly altered in VTA and SNc ($P > 0.05$, $n = 9$). On the right, bar graphs plot the abundance of *Ucp4* and *Ucp5* mRNAs, normalized by that of *Ucp2* in each sample. The relative abundance of *Ucp4* and *Ucp5* mRNA was decreased in the SNc of *DJ-1* knockout mice ($P < 0.05$, $n = 9$). *Ucp4* mRNA abundance was higher in VTA from *DJ-1* knockout mice ($P < 0.05$, $n = 9$), whereas *Ucp5* mRNA was unchanged. Error bars are s.e.m. **d**, Summary of the presented results linking calcium entry through L-type channels during pacemaking with elevated mitochondrial oxidant stress and opening of UCPs. The model also proposes that oxidized DJ-1 translocates to the nucleus and increases the transcription of *Ucp4* and *Ucp5* or the stability of the resulting mRNAs, leading to increased concentrations of UCP in the IMM. OXPHOS, oxidative phosphorylation; ROS, reactive oxygen species.

mice showing strong mito-roGFP expression. In the presence of a strong reducing agent as an estimate of roGFP concentration, we inferred that there was no significant difference between the expression levels of SNc and VTA neurons (Supplementary Fig. 7). The oxidation state of mitochondria was also verified in a second line of mice, one with lower mito-roGFP expression levels (Supplementary Fig. 7), and in cultured cells expressing mito-roGFP (Supplementary Fig. 6). The single-cell PCR approach that we used has been described previously[4]. Relative gene expression of UCPs was performed by reverse transcriptase reaction followed by quantitative PCR analysis. The immunostaining of manganese superoxide dismutase used standard approaches. Sample n represents number of mice. Our statistical analysis used non-parametric tests with a threshold for significance of less than 0.05.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Albin, R. L., Young, A. B. & Penney, J. B. The functional anatomy of disorders of the basal ganglia. *Trends Neurosci.* **18,** 63–64 (1995).
2. Schapira, A. H. Mitochondria in the aetiology and pathogenesis of Parkinson's disease. *Lancet Neurol.* **7,** 97–109 (2008).
3. Puopolo, M., Raviola, E. & Bean, B. P. Roles of subthreshold calcium current and sodium current in spontaneous firing of mouse midbrain dopamine neurons. *J. Neurosci.* **27,** 645–656 (2007).
4. Chan, C. S. et al. 'Rejuvenation' protects neurons in mouse models of Parkinson's disease. *Nature* **447,** 1081–1086 (2007).
5. Khaliq, Z. M. & Bean, B. P. Pacemaking in dopaminergic ventral tegmental area neurons: depolarizing drive from background and voltage-dependent sodium conductances. *J. Neurosci.* **30,** 7401–7413 (2010).
6. Guzman, J. N., Sanchez-Padilla, J., Chan, C. S. & Surmeier, D. J. Robust pacemaking in substantia nigra dopaminergic neurons. *J. Neurosci.* **29,** 11011–11019 (2009).
7. Dooley, C. T. et al. Imaging dynamic redox changes in mammalian cells with green fluorescent protein indicators. *J. Biol. Chem.* **279,** 22284–22293 (2004).
8. Matlib, M. A. et al. Oxygen-bridged dinuclear ruthenium amine complex specifically inhibits Ca^{2+} uptake into mitochondria *in vitro* and in situ in single cardiac myocytes. *J. Biol. Chem.* **273,** 10223–10231 (1998).
9. Nicholls, D. G. & Ferguson, S. J. *Bioenergetics 3* (Academic, 2002).
10. Kahle, P. J., Waak, J. & Gasser, T. DJ-1 and prevention of oxidative stress in Parkinson's disease and other age-related disorders. *Free Radic. Biol. Med.* **47,** 1354–1361 (2009).
11. Ehrenberg, B., Montana, V., Wei, M. D., Wuskell, J. P. & Loew, L. M. Membrane potential can be determined in individual cells from the Nernstian distribution of cationic dyes. *Biophys. J.* **53,** 785–794 (1988).
12. Rasola, A. & Bernardi, P. The mitochondrial permeability transition pore and its involvement in cell death and in disease pathogenesis. *Apoptosis* **12,** 815–833 (2007).
13. Krauss, S., Zhang, C. Y. & Lowell, B. B. The mitochondrial uncoupling-protein homologues. *Nature Rev. Mol. Cell Biol.* **6,** 248–261 (2005).
14. Brand, M. D. et al. Mitochondrial superoxide and aging: uncoupling-protein activity and superoxide production. *Biochem. Soc. Symp.* **71,** 203–213 (2004).
15. Andrews, Z. B. et al. Uncoupling protein-2 is critical for nigral dopamine cell survival in a mouse model of Parkinson's disease. *J. Neurosci.* **25,** 184–191 (2005).
16. Zhang, C. Y. et al. Genipin inhibits UCP2-mediated proton leak and acutely reverses obesity- and high glucose-induced beta cell dysfunction in isolated pancreatic islets. *Cell Metab.* **3,** 417–427 (2006).
17. Echtay, K. S. et al. Superoxide activates mitochondrial uncoupling proteins. *Nature* **415,** 96–99 (2002).
18. Papa, S. & Skulachev, V. P. Reactive oxygen species, mitochondria, apoptosis and aging. *Mol. Cell. Biochem.* **174,** 305–319 (1997).
19. Canet-Aviles, R. M. et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc. Natl Acad. Sci. USA* **101,** 9103–9108 (2004).
20. Bonifati, V. et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset Parkinsonism. *Science* **299,** 256–259 (2003).
21. Cookson, M. R. DJ-1, PINK1, and their effects on mitochondrial pathways. *Mov. Disord.* **25** (suppl. 1), S44–S48 (2010).
22. Bender, A. et al. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nature Genet.* **38,** 515–517 (2006).
23. Bender, A. et al. Dopaminergic midbrain neurons are the prime target for mitochondrial DNA deletions. *J. Neurol.* **255,** 1231–1235 (2008).
24. Krishnan, K. J., Greaves, L. C., Reeve, A. K. & Turnbull, D. M. Mitochondrial DNA mutations and aging. *Ann. NY Acad. Sci.* **1100,** 227–240 (2007).
25. Nicholls, D. G. Oxidative stress and energy crises in neuronal dysfunction. *Ann. NY Acad. Sci.* **1147,** 53–60 (2008).
26. Eisenberg, M. J., Brox, A. & Bestawros, A. N. Calcium channel blockers: an update. *Am. J. Med.* **116,** 35–43 (2004).
27. Becker, C., Jick, S. S. & Meier, C. R. Use of antihypertensives and the risk of Parkinson disease. *Neurology* **70,** 1438–1444 (2008).
28. Ritz, B. et al. L-type calcium channel blockers and Parkinson disease in Denmark. *Ann. Neurol.* **67,** 600–606, doi:10.1002/ana.21937 (2010).

Author Contributions D.J.S. was responsible for the overall direction of the experiments, analysis of data, construction of figures and communication of the results. J.N.G. and J.S.-P. were responsible for the design and execution of experiments, as well as the analysis of results. D.W. provided expertise on optical approaches. E.I. conducted the immunocytochemical experiments. P.T.S. and J.K. were responsible for the generation of the TH-mito-roGFP mice; they also participated in the design, analysis and communication of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J.S. (j-surmeier@northwestern.edu).

METHODS

Tissue preparation. Acute midbrain coronal slices (220-μm thick) were obtained from male mice between postnatal ages 21 and 30 d, unless specified otherwise. The handling of mice and all procedures performed on them were approved by the Northwestern University Animal Care and Use Committee and were in accordance with the National Institutes of Health Guide to the Care and Use of Laboratory Animals and Society for Neuroscience guidelines. Mice were anaesthetized with a ketamine–xylazine mixture followed by transcardial perfusion with ice-cold, oxygenated artificial cerebrospinal fluid (ACSF) containing 125 mM NaCl, 2.5 mM KCl, 25 mM NaHCO$_3$, 1.25 mM NaH$_2$PO$_4$, 2 mM CaCl$_2$, 1 mM MgCl$_2$ and 25 mM dextrose, pH 7.3, osmolality of 315–320 mOsM. After perfusion, mice were decapitated and brains were removed rapidly, followed by sectioning in ice-cold oxygenated ACSF using a vibratome (VT1000S Leica Microsystems). Midbrain slices were kept at 34 °C for 30 min before 2PLSM imaging experiments were performed. Dopaminergic neurons from the ventral tier of the SNc were identified during the imaging experiments on the basis of anatomical location in the slice and cell body size, and VTA neurons were identified by their anatomical location medial to the SNc.

Generation of transgenic mice with mitochondrially targeted roGFP in dopaminergic neurons. Mitochondrial redox status was assessed with roGFP. Oxidation of two cysteine thiols located at the outer surface of the protein shifts the excitation spectrum of the protein, allowing ratiometric, real-time visualization of oxidation state[7]. roGFP plasmid was generated from the pEGFP-N1 plasmid (CLONTECH) by substituting surface-exposed residues on enhanced GFP with cysteines in appropriate positions as previously described, using a QuickChange site-directed mutagenesis kit (Stratagene). To target roGFP to the mitochondrial matrix, the matrix targeting sequence (MTS) from cytochrome oxidase subunit 4 was added at the amino terminus using the primer pair 5′-AATGCCGCTAGCGCCACCATGCTGAGCCTGCGCCAGAGCATCCGCT TCTTCAAGCGCCAGGCGGCATCATGGTGAGCAAGGGCGAGGAGCTG-3′ and 5′-CCGGTCGAGTTACTTGTACAGCTCGTCCATGCC-3′ while adding restriction sites for NheI and XhoI. The resulting PCR product was cloned into the pcDNA 3.1(+) vector (Invitrogen). MTS–roGFP was again amplified using primer pair 5′-AATGCCGATATCGCCACCATGCTGAGC-3′ and 5′-CCGCTCGAGT TACTTGTACAGCTCGTCCATGCC-3′, adding restriction sites for EcoRV and XhoI. PCR products were cloned in pCR4-TOPO vector (Invitrogen) and excised with EcoRI.

To drive expression in dopaminergic neurons, this construct was ligated into the TH-SK3-PA plasmid containing the tyrosine hydroxylase (TH) promoter, a 9.0-kilobase (kb) genomic fragment, located 5′ to the EcoRI restriction site. The TH-SK3-PA vector was kindly provided by Dr X. Zhuang at the University of Chicago, with permission from Dr J. H. Son[29] (Laboratory of Molecular Neurobiology, The W. M. Burke Medical Research Institute, White Plains, New York). Correct orientation of the gene was verified by sequencing. To allow *in vivo* expression of MTS–roGFP in dopaminergic neurons, a 10.5-kb SalI–HindIII fragment containing tyrosine hydroxylase promoter/MTS–roGFP/polyadenylation sequences was purified by electrophoresis with 0.8% agarose gel, and eluted with an Elutip column as per the manufacturer's instructions (Whatman Schleicher & Schuell). The purified DNA was used to generate transgenic mice by pronuclear microinjection at the Northwestern University Transgenic and Targeted Mutagenesis core. Several founder transgenic mice were identified by PCR using primers that hybridize with enhanced GFP. The founders for the roGFP mice were generated in a B6SJL genetic background and backcrossed to C57BL6 for our imaging experiments.

Generation of transgenic mice carrying the CMV-promoter-driven, mitochondrially targeted roGFP gene. The oxidation and reduction status of the mitochondrial matrix was assessed by targeting the expression of a redox-sensitive protein (roGFP) to that compartment. A roGFP plasmid was generated from the pEGFP-N1 plasmid (CLONTECH) by site-directed mutagenesis using the Stratagene QuickChange kit. To target roGFP to mitochondria, MTS was added at the amino terminus using the primer pair 5′-AATGCC<u>GCTAGC</u>GCCA CC*ATGCTGAGCCTGCGCCAGAGCATCCGCCTTCTTCAAGCGCAGGCGGCAT Catggtgagcaagggcgaggagctg*-3′ and 5′-CCG<u>CTCGAG</u>TTACTTGTACAGCTC GTCCATGCC-3′ while adding restriction sites for NheI and Xho I (underlined). The resultant plasmid was cloned into pcDNA 3.1(+) vector (Invitrogen). To drive mitochondrially targeted roGFP expression in different regions of the brain, a 2.05-kb MluI–DraIII fragment containing CMV promoter/MTS–roGFP/BGH polyadenylation sequences was purified and used in pronuclear injections at the Northwestern University Transgenic and Targeted Mutagenesis core. Several founder transgenic mice were identified by PCR using enhanced GFP primers.

2PLSM: calcium imaging. For dendritic measurements, SNc dopaminergic neurons in tissue slices (as described above) were loaded with Alexa Fluor 594 (50 μM)

and Fluo-4 (200 μM) through the patch pipette. All experiments were performed at 34–35 °C. Images were acquired with a ×60, NA-0.9 water-immersion lens. Dyes were allowed to equilibrate for at least 15 min before imaging. The two-photon excitation source was a Chameleon Ultra 2 tunable laser system (680–1,080 nm) using a Ti:sapphire gain medium with all-solid-state active components and a computer-optimized algorithm to ensure reproducible excitation wavelength, average power and peak power (Coherent Laser Group). Optical signals were acquired using an 810-nm excitation beam (80-MHz pulse repetition frequency and ~250-fs pulse duration) to excite Alexa and Fluo-4 dyes simultaneously. Laser power attenuation was achieved with two Pockels-cell electro-optic modulators (model 350-80, Con Optics). The two cells were aligned in series to provide enhanced modulation range for fine control of the excitation dose (0.1% steps over four decades). The laser-scanned images were acquired with a Bio-Rad Radiance MPD system. The fluorescence emission was collected with non-descanned photomultiplier tubes (PMTs). The green fluorescence (500–550 nm) was detected by a bialkali-cathode PMT and the red fluorescence (580–640 nm) was collected by a multi-alkali-cathode (S-20) PMT. The system digitizes the current from detected photons to 12 bits. The laser light transmitted through the sample was collected by the condenser lens and sent to another PMT to provide a bright-field transmission image in registration with the fluorescent images. Measurements were taken in a sample plane along dendritic segments (100–150 mm from the soma). Line scan signals were acquired (as described above) at 6 ms per line and 512 pixels per line with 0.18-μm pixels and 10-μs pixel dwell time. The time between the control and isradipine treatment measurements was 5–7 min.

2PLSM: TMRM mitochondrial imaging. Brain slices from wild-type or *DJ-1* knockout mice (mixed background C57BL6x129SvEv) were incubated in 2–4 μM TMRM for 30 min at 34–35 °C; excess dye was then washed out with a TMRM-free ACSF solution. Imaging experiments were performed with TMRM-free ACSF solution at physiological temperatures (34–35 °C). Fluorescence (550–640 nm) was collected by a multi-alkali-cathode (S-20) PMT and a Dodt contrast-detector system that provided a bright-field transmission image (Prairie Technologies). The two-photon excitation source was a Chameleon Ultra 2 tunable laser system (706–1,000 nm) using a Ti:sapphire gain medium with all-solid-state active components and a computer-optimized algorithm to ensure reproducible excitation wavelength, average power and peak power (Coherent Laser Group). Signals were acquired using an 830-nm excitation beam (80-MHz pulse repetition frequency and ~250-fs pulse duration), in a fixed plane of focus with a pixel size between 0.18 and 0.21 μm. Fluorescence measurements in a region of interest (ROI) were monitored to ensure stationarity of the signal; samples with a drifting baseline (due to photobleaching of dye and/or washout) were discarded. Time-series scanning (1,000 frames) in a fixed plane was performed with a 4-μs dwell time at a rate of 2.5 frames per second. Drug applications ranged in duration from 30 to 60 min with similar control periods for comparison. Four to five ROIs in the cell body and one ROI in the nucleus were monitored in a typical experiment, and changes in TMRM fluorescence were plotted as a function of time.

The flickering frequency was determined by counting the number of transitions in 100-s epochs; fluorescence histograms were calculated with MATLAB (release 2009b). The change in mitochondrial membrane potential during flickering was estimated from the fluorescence in an ROI using a Nernst equation, $V = (RT/zF)\ln(\beta F_m/F_n)$, where R is the gas constant, T is temperature, F is Faraday's constant, $z = 1$, F_m is the fluorescence in the mitochondrial ROI, F_n is the fluorescence of the nucleus in the same optical plane and β is a scaling factor. The scaling factor was calculated by assuming that the mitochondrial membrane potential was −150 mV when the TMRM fluorescence was maximal. The existence of quenching was checked by application of the ATP synthase inhibitor oligomycin (10 μM); invariably, maximal fluorescence remained near that seen before application of oligomycin. For each flickering event, the percentage change in V was calculated by taking the difference in the estimates before and during the drop in fluorescence; the fluorescence during the drops was averaged.

These studies indicated that there was little or no quenching at the TMRM concentrations used; moreover, they excluded a role for ATP synthase in the origin of the flickering. These results also suggest that the TMRM concentration achieved in the mitochondria was not high enough to make fluorescence quenching an issue. Experiments illustrated in Supplementary Fig. 1 were performed using the internal solution described for 2PLSM calcium imaging, except that the internal solution contained 400 nM TMRM instead of Fluo-4. Fluorescence imaging was performed as above; the electrical activity of the cell was controlled in current-clamp mode as described in the figure legend.

2PLSM: mitochondrial roGFP imaging. Midbrain slices from wild-type, roGFP transgenic mice or *DJ-1* knockout mice (crossed with roGFP transgenic mice) were incubated in ACSF at physiological temperatures (34–35 °C) followed by a recovery period at room temperature. Optical imaging of roGFP signals acquired using a

920-nm excitation beam (80-MHz pulse repetition frequency and ~250-fs pulse duration), in a fixed plane of focus with a pixel size between 0.18 and 0.21 μm and a 2–3-μs pixel dwell time. This wavelength was chosen by determining the two-photon excitation spectrum of mito-roGFP in cultured dopaminergic neurons from TH-mito-roGFP mice. The two-photon excitation source was a Chameleon Ultra 2 tunable laser system (706–1,000 nm) using a Ti:sapphire gain medium with all-solid-state active components and a computer-optimized algorithm to ensure reproducible excitation wavelength, average power and peak power (Coherent Laser Group). The roGFP fluorescence (490–560 nm) was detected by a bialkali-cathode PMT and a Dodt contrast-detector system that provided a bright-field transmission image (Prairie Technologies). Sixty frames of the roGFP signal were collected in one optical plane at a rate of 3–4 frames per second. Experiments were performed at room temperature (20–22 °C) or near-physiological temperatures (34–35 °C). Records with a drifting baseline (due to photobleaching or photo-oxidation of roGFP) were discarded. At the end of all experiments, the maximum and minimum fluorescences of mito-roGFP were determined by application of 2 mM dithiothreitol (DTT) to reduce the mitochondria fully, and then 100 μM Aldrithiol (Ald) to oxidize the mitochondria fully. The relative oxidation was then calculated as $1 - [(F - F_{Ald})/(F_{DTT} - F_{Ald})]$.

Quantitative polymerase chain reaction. Brain regions (midbrain: SNc and VTA; cerebellum; cortex; hippocampus; striatum) were micro-dissected from wild-type or *DJ-1* knockout mice (ages P28–32) and total RNA was extracted using Trizol reagent (Invitrogen); complementary DNA was generated by reverse transcription (Quanta Biosciences) from both knockout and wild-type samples. cDNA from each sample was analysed using quantitative PCR (qPCR)[30]. qPCR experiments were performed using Sybr-Green and sense and antisense primers (Integrated DNA Technologies): primers for UCP2, UCP4 and UCP5 were reported previously[15]; primers for the other cDNAs were manganese superoxide dismutase (MnSOD) (sense, 5′-ACAACTCAGGTCGCTCTTCAG-3′; antisense, 5′-GATAGCCTCCAGCAACTCTCC-3′), catalase (sense, 5′-GCCAGCGACC AGATGAAG-3′; antisense, 5′-AAGTTTATCTCCTATTGGGTTCC-3′) and glutathione peroxidase 1 (sense, 5′-CGAAGTGAATGGTGAGAAGG-3′; antisense, 5′-TCAAAGTTCCAGGCAATGTC-3′). qPCR cycling parameters were as follows: 1 cycle at 95 °C for 3 min, followed by 40 cycles each consisting of 15 s at 94 °C, 1 min at 60 °C and 30 s at 72 °C, followed Sybr-Green fluorescence reading. After the completion of 40 cycles, we analysed the melting temperature curve of each sample. Expression level was estimated using the comparative C_T approximation method to calculate the relative abundance of mRNA. Samples were run in triplicate for accuracy and normalized to expression levels of UCP2 to obtain the ΔC_T ($\Delta C_T = C_T(UCP4/5) - C_T(UCP2)$). Paired ΔC_Ts for the wild-type and *DJ-1* knockout groups and the mRNA abundance were calculated from the equation $2^{-\Delta\Delta C_T}$ (ref. 31).

MnSOD immunostaining. Wild-type and *DJ-1* knockout mice ($n = 3$ per genotype) were perfused transcardially with 4% paraformaldehyde in 0.1 M phosphate buffer, pH 7.3. Brains were removed and cryoprotected in 30% sucrose solution, and thin brain sections (30-μm thick) were obtained using a freezing microtome.

Sections were blocked in 5% normal goat serum followed by incubation with a polyclonal antibody against MnSOD (16–18 h at 4 °C, working dilution of 1:300; Stressgen). Staining was visualized with a goat anti-rabbit secondary antibody conjugated to Alexa 594 (2-h incubation at room temperature, working dilution of 1:300) (Invitrogen). For cultured SNc dopaminergic neurons, ventral midbrain regions of neonatal TH-mito-roGFP mice (P0–3) were processed as described previously[32,33]. Cultures were fixed with 4% paraformaldehyde followed by immunostaining using the same conditions described above for labelling MnSOD in brain sections. Images were acquired with a laser-scanning confocal microscope (LSM510, Zeiss) and stored at 12-bit image depth at a resolution of $1,024 \times 1,024$ pixels. The contrast range in images of tissue from wild-type and *DJ-1* knockout mice was the same.

Statistical analysis. Imaging data collected was analysed with IGOR PRO 6.0 (Wavemetrics), MATLAB (Mathworks) or GRAPHPAD PRISM 5.0 (GraphPad Software). The stimulation, display and analysis software for the two-photon imaging data was analysed using a custom-written shareware package, WINFLUOR, PICVIEWER and POWERCAL, kindly provided by J. Dempster (Strathclyde University, Glasgow, UK). Data were summarized using box plots showing median values and interquartile range to represent statistical dispersion for small sample sizes. Sample *n* typically represents the number of mice used for experiments. For the qPCR experiments, data represents mean ± s.e.m. Statistical analysis was done with SIGMASTAT 3.5 (Systat Software) using non-parametric tests (Mann–Whitney rank-sum test for comparing between two groups; Kruskal–Wallis analysis of variance with Dunnet's *post hoc* test for comparison of multiple groups). Paired statistical analyses were performed using the Wilcoxon signed-rank test. Probability threshold for statistical significance was $P < 0.05$.

Pharmacological reagents and channel ligands. Reagents were purchased from Sigma except for isradipine, DTT (Tocris-Cookson), TMRM and Fura-2 (Invitrogen/Molecular Probes), and genipin (Sigma and Wako Reagents). Drug stock solutions were prepared in deionized water, DMSO, ethanol or methanol, as specified by the manufacturer's instructions, on the day of experiment. Stocks were diluted to final concentrations in ACSF to achieve a final solvent concentration of less than 0.1% v/v.

29. Son, J. H. *et al.* Neuroprotection and neuronal differentiation studies using substantia nigra dopaminergic cells derived from transgenic mouse embryos. *J. Neurosci.* **19,** 10–20 (1999).
30. Bookout, A. L., Cummins, C. L., Mangelsdorf, D. J., Pesola, J. M. & Kramer, M. F. High-throughput real-time quantitative reverse transcription PCR. *Curr. Protoc. Mol. Biol.* Chapter 15, Unit 15 18 (2006).
31. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3,** 1101–1108 (2008).
32. Tian, X., Kai, L., Hockberger, P. E., Wokosin, D. L. & Surmeier, D. J. MEF-2 regulates activity-dependent spine loss in striatopallidal medium spiny neurons. *Mol. Cell. Neurosci.* **44,** 94–108 (2010).
33. Fath, T., Ke, Y. D., Gunning, P., Gotz, J. & Ittner, L. M. Primary support cultures of hippocampal and substantia nigra neurons. *Nature Protocols* **4,** 78–85 (2009).

LETTER

Planar polarized actomyosin contractile flows control epithelial junction remodelling

Matteo Rauzi[1]†, Pierre-François Lenne[1] & Thomas Lecuit[1]

Force generation by Myosin-II motors on actin filaments drives cell and tissue morphogenesis[1-15]. In epithelia, contractile forces are resisted at apical junctions by adhesive forces dependent on E-cadherin[16], which also transmits tension[6,17-19]. During *Drosophila* embryonic germband extension, tissue elongation is driven by cell intercalation[20], which requires an irreversible and planar polarized remodelling of epithelial cell junctions[4,5]. We investigate how cell deformations emerge from the interplay between force generation and cortical force transmission during this remodelling in *Drosophila melanogaster*. The shrinkage of dorsal–ventral-oriented ('vertical') junctions during this process is known to require planar polarized junctional contractility by Myosin II (refs 4, 5, 7, 12). Here we show that this shrinkage is not produced by junctional Myosin II itself, but by the polarized flow of medial actomyosin pulses towards 'vertical' junctions. This anisotropic flow is oriented by the planar polarized distribution of E-cadherin complexes, in that medial Myosin II flows towards 'vertical' junctions, which have relatively less E-cadherin than transverse junctions. Our evidence suggests that the medial flow pattern reflects equilibrium properties of force transmission and coupling to E-cadherin by α-Catenin. Thus, epithelial morphogenesis is not properly reflected by Myosin II steady state distribution but by polarized contractile actomyosin flows that emerge from interactions between E-cadherin and actomyosin networks.

The planar polarized remodelling of cell junctions[4,5] that occurs during germband extension (GBE) is shown in Fig. 1a. Myosin II (Myo-II) is concentrated in 'vertical' junctions[4,21] and directs junction shrinkage by increasing junctional tension[7,12]. To understand how Myo-II planar polarity is established, we investigated changes in Myo-II distribution at the onset of GBE. We used a fusion between Myo-II regulatory light chain (MRLC, called Sqh in *Drosophila*) and Cherry (MRLC–Cherry)[15] together with E-cad–GFP to mark adherens junctions (AJs). When the epithelium is formed, MRLC–Cherry is visible in aggregates in the medial region of AJs (Fig. 1b). Subsequently, MRLC–Cherry is also detected at the cortex of AJs of intercalating cells (Fig. 1b). An MRLC–GFP fusion rescuing a null *sqh*[AX3] mutant (Fig. 1c) and an antibody against endogenous Myo-II heavy chain (not shown) displayed the same features. Thus two Myo-II populations exist during cell intercalation: a medial and a junctional pool (Supplementary Fig. 1). Labelling of F-actin with Utrophin–GFP (Utr–GFP) shows a network spanning the AJs (Fig. 1d, Supplementary Fig. 1, Supplementary Movie 1a). This network is thin (<500 nm), and contains filaments at low density (mesh size 0.5–2 μm) that overlap and intersect in the form of brighter puncta, which are more apparent in a slightly less apical focal plane intersecting the AJs (Supplementary Movie 1b, Supplementary Fig. 1). Thus, both Myo-II pools are part of a large-scale actomyosin network, spanning multiple cells, which contrasts with previous descriptions focused on junctional actin and Myo-II (refs 4, 5, 7, 12, 18, 21–23).

Live imaging of Utr–GFP and MRLC–GFP indicated complex dynamics (Supplementary Movies 1a, b and 2). The F-actin mesh fluctuated, with the mesh changing size in a few tens of seconds (Fig. 1d and Supplementary Movie 1a, b). Myo-II formed small clusters

(presumably Myo-II minifilaments), which coalesced into large (~1 μm) medial aggregates on similar timescales (Fig. 1e, Supplementary Movie 2). Co-imaging of Utr–GFP and MRCL–Cherry revealed
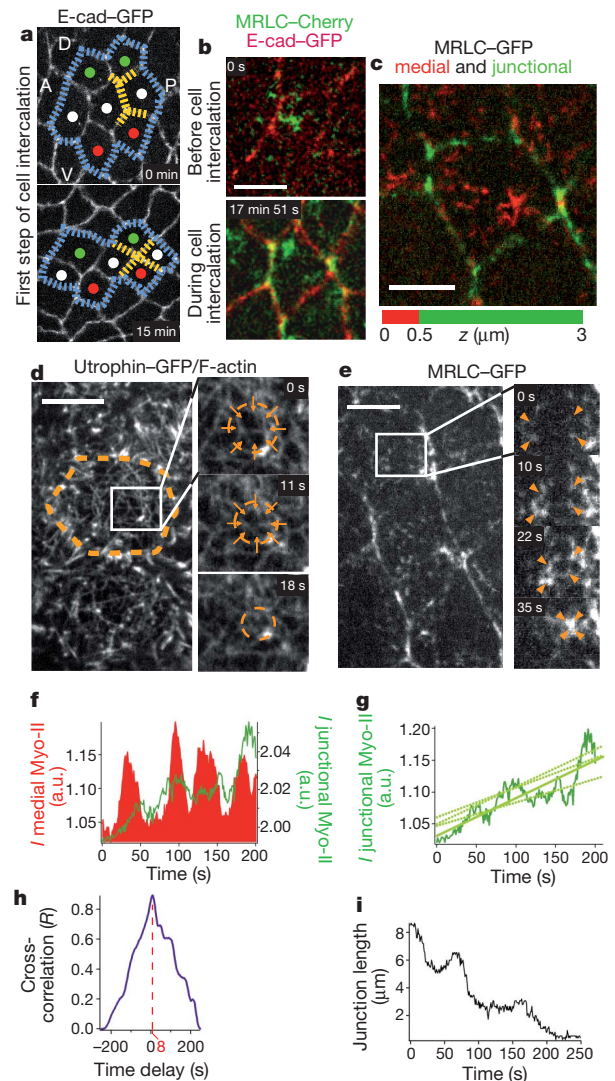
Figure 1 | Two pulsating pools of acto-myosin in intercalating cells.
a, Polarized junction shrinkage during cell intercalation. A, P, D and V denote respectively dorsal, ventral, anterior and posterior. **b**, Localization of Myo-II and E-cad before and during intercalation. **c**, Respective distribution of medial (red) and junctional (green) Myo-II along the apico-basal (z) axis. **d**, **e**, Apical F-actin coalesces locally (**d**, magnified in right panels, arrows), while medial Myo-II clusters (**e**, magnified in right panels, arrowheads). **f**, Myo-II pulses in the medial (red) and junctional (green) regions. **g**, Average junctional Myo-II (dark green) and linear fits for different junctions. **h**, Temporal cross-correlation of the curves in **f**. R is the correlation coefficient. **i**, Evolution of junctional length. Scale bars, 5 μm.

[1]IBDML, UMR6216 CNRS-Université de la Méditerranée, Campus de Luminy, case 907, 13288 Marseille Cedex 09, France. †Present address: EMBL, Meyerhofstrasse 1, 69126 Heidelberg, Germany.

that actin and Myo-II coalesce together during aggregation (Supplementary Fig. 2a, Supplementary Movie 3; Methods), reflecting local and transient contractions within the actomyosin network, as also reported in the *Drosophila* mesoderm and the one-cell stage *Caenorhabditis elegans* cortex[15,24].

To further investigate the functions of the medial and junctional Myo-II networks, we monitored their temporal evolution during intercalation (Fig. 1f–i). Both medial and junctional Myo-II, respectively in the vicinity of and at shrinking junctions, fluctuated in intensity (Fig. 1f). In addition to being pulsed, the intensity of junctional Myo-II gradually increased (Fig. 1g). Meanwhile, the changes in 'vertical' junction length are irregular, showing successive steps of shrinkage and arrest (Fig. 1i). In some cases, however, transient relaxation was observed (17.6%, $N = 17$).

To disentangle this complex behaviour and relate contractile dynamics of medial and junctional networks with junction shrinkage, we conducted temporal cross-correlation of fluorescence intensity (Online Methods). Correlation between temporal profiles of MRLC–GFP intensity at the junctions and in the medial regions is high (mean $<R> = 0.86$, Fig. 1h), indicative of similar overall dynamics. Moreover, medial pulses precede junctional pulses by 8 ± 4 s (mean $\pm$ s.d. hereafter, Fig. 1h and Supplementary Fig. 3).

We then compared rates of junction shrinkage with rates of MRLC–GFP intensity changes (Fig. 2a left), which correspond to local accumulations of Myo-II by contraction (Fig. 1e). The maximum of the MRLC–GFP contraction rate in the medial region precedes that of junctional MRLC–GFP by an average of 10.5 ± 2.5 s (Fig. 2b left). Thus contraction of Myo-II occurs in the medial region first and subsequently at junctions (Fig. 2b left, right). Each step of junction shrinkage was associated with

tandem medial and junctional Myo-II pulses (Fig. 2a right, horizontal braces). Temporal cross-correlation indicated that the peak rate of junction shrinkage precedes that of junctional Myo-II accumulation by 9 ± 3 s (Fig. 2b left), indicating that junctional Myo-II accumulation cannot cause the shrinkage steps. However, peak junction shrinkage rate temporally coincided with the peak rate of medial Myo-II contraction (Fig. 2b left, right) suggesting a mechanical contribution to shrinkage increments.

To test this, we used laser nanodissection[12] to locally disrupt medial Myo-II clusters at the vicinity of shrinking junctions. Each ablation pulse produced a collapse of the Myo-II pulse and a transient and reversible relaxation of junction length without affecting junctional Myo-II (Fig. 2c left, right, Supplementary Movie 4). Thus, medial Myo-II mechanically causes junction shrinkage. This led us to investigate the function of junctional Myo-II pulses, as previous studies showed it was essential for global junction shrinkage[4,7,12]. Close inspection reveals two situations: (1) in most cases (88%, $N = 17$), medial Myo-II pulses are followed by junctional pulses, and shrinkage steps proceed successfully without relaxation (14/15 cases, Fig. 2a right); (2) occasionally, (12%) medial pulses are not followed by junctional pulses and shrinkage steps relax in all cases (Fig. 2d). Relaxation correlates with failure to sustain junctional Myo-II and with an overall decrease of Myo-II at junctions (Fig. 2d left, right). This suggests that junctional Myo-II stabilizes junction length.

Together these observations point to a mechanical 'division of labour', where medial Myo-II pulses shrink, and sustained junctional Myo-II accumulation stabilizes, junction length. This iterative cycle ensures persistent shrinkage.
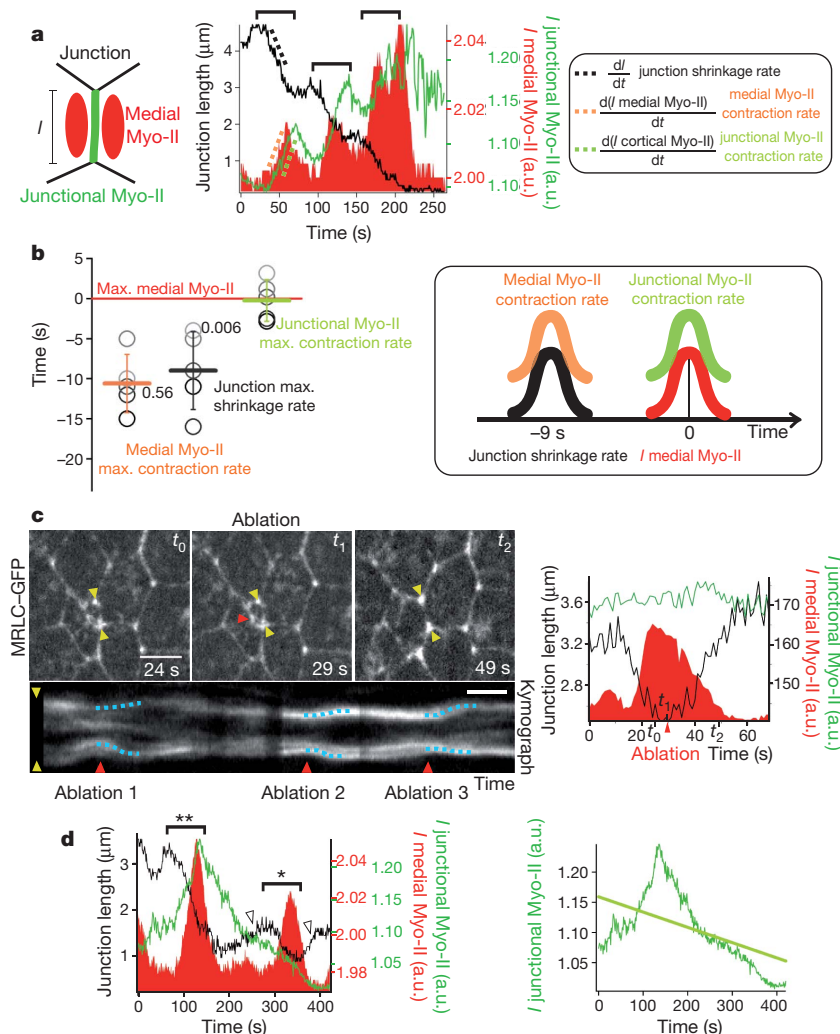
Figure 2 | Medial and junctional Myo-II pools have different mechanical roles. a, Left: cartoon depicting a vertical junction (length l) and regions where medial and junctional Myo-II are measured. Right: evolution of junction length and Myo-II intensities (I). Brackets show clusters of events and dashed lines represent the rates of changes. **b**, Time delays using medial Myo-II intensity maximum as a reference. Mean and s.d. are shown in the left panel. **c**, Left: evolution of a junction (between yellow arrowheads) before (t_0), during (t_1) and after (t_2) focal ablation of a Myo-II pulse (red arrowhead). Bottom, kymograph of the same junction with three ablation events (red arrowheads). Blue dashed lines mark junction relaxation. Scale bar, 25 s. Right: length of junction (from **c**, left) and medial and junctional Myo-II intensity as a function of time. The red arrowhead marks the ablation. **d**, Left: junction relaxation (arrowheads) when a medial pulse is not followed by a junctional pulse (*) or when junctional Myo-II is not sustained (**). Right: junctional Myo-II intensity (dark green) and linear fit (light green). Scale bars in **c**, 5 µm.

These observations suggested that both processes may be spatially coordinated. Indeed, medial pulses show a planar polarized distribution like junctional Myo-II. Defining four quadrants (anterior, A; posterior, P; dorsal, D; and ventral, V; diagram in Fig. 3a right) in the medial region of cells, we determined the integrated intensity ratio of (A+P)/(D+V) MRLC–GFP in time series (Fig. 3a left). Intercalating germband cells exhibit a significant medial Myo-II polarity compared to non-intercalating head cells or to germband cells of *Krüppel*(*Kr*) RNAi embryos where planar cell polarization is affected[4,20] (Fig. 3a left, Supplementary Movie 6).

We next investigated the spatial dynamics of medial and junctional actomyosin networks. Co-imaging of Utr–GFP and MRLC–Cherry and particle imaging velocimetry (PIV) indicated that F-actin and Myo-II have very similar dynamics and that actomyosin clusters flow in the plane of the medial region (Fig. 3b, Supplementary Movie 3). Myo-II was moving slightly (22%) but consistently faster than F-actin (Supplementary Fig. 5), in agreement with the idea that Myo-II is responsible for flow.
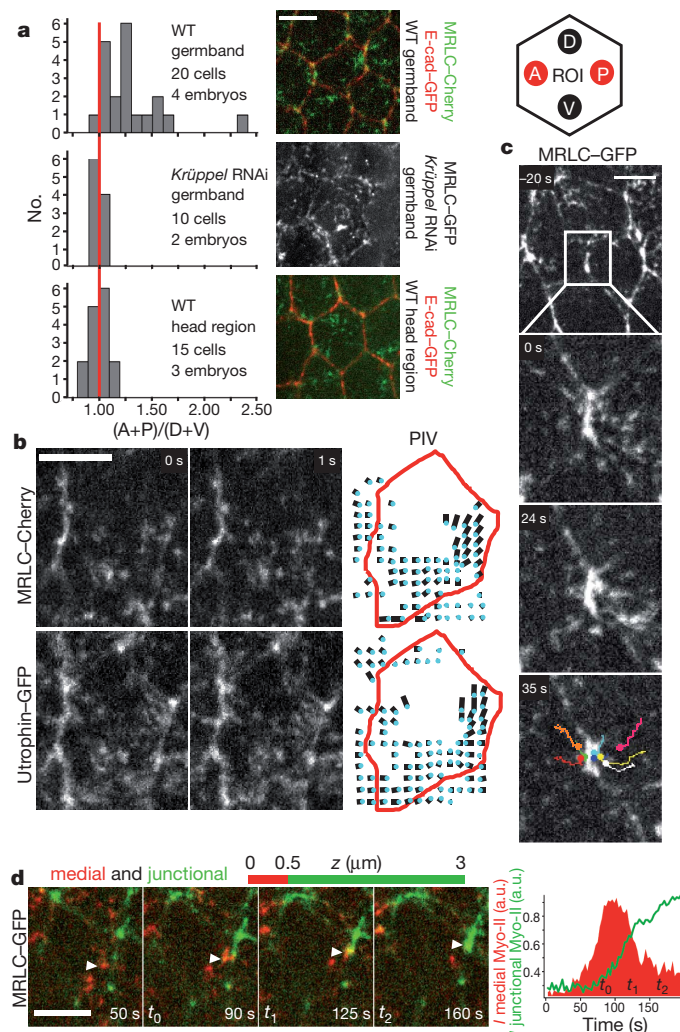
Figure 3 | Medial Myosin-II displays anisotropic flow and feeds 'vertical' junctions. a, Left: histograms of medial Myo-II relative intensities in (A+P) regions over (D+V) regions (see diagram at right for nomenclature) in the germband of wild-type (WT) and *Krüppel* (*Kr*) RNAi embryos, and the head of WT embryos. WT/*Kr*RNAi: $P = 0.0007$, WT(germband)/WT(head): $P = 0.001$ (T-student). Right; representative images of cells with MRCL–Cherry and E-cad–GFP. **b**, Comparative PIV of Utr–GFP and MRLC–Cherry in a cell outlined in red. Blue dots mark vector tips. **c**, Medial Myo-II flowing to a vertical junction. Tracking of speckles is showed in coloured lines (right). **d**, Left: a medial cluster (red, arrowhead) flows and fuses to the junctional Myo-II pool (green); right: corresponding quantification. Scale bars, 5 μm.

Tracking of Myo-II speckles (Fig. 3c, Supplementary Movies 2, 5) or of F-actin with Myo-II (Supplementary Fig. 2b) indicated that the polarized distribution of medial Myo-II results from the lateral flow of medial pulses towards 'vertical' junctions. In *Kr*RNAi embryos, this movement occurred randomly (Supplementary Movie 6), consistent with the loss of medial Myo-II polarity (Fig. 3a left).

The polarized flow of Myo-II ($0.11 \pm 0.03 \, \mu m \, s^{-1}$) could either reflect a movement of Myo-II minifilaments or the propagation of contractile waves. We tested these alternatives by photobleaching medial MRLC–GFP clusters. The fluorescence recovery in the bleached area (recovery fractions $34 \pm 10\%$ ($N = 5$), $\tau_{1/2} = 4 \pm 1 \, s$, Supplementary Fig. 6a, b, d) was low compared to the junctions (recovery fractions ~70% (ref. 8), not shown). Moreover, no new cluster appears in the vicinity of bleached pulses, as would be expected for contractile waves (Supplementary Fig. 6a). Together this indicates that medial flows correspond to the movement of relatively stable Myo-II filaments. Fluorescence recovery after photobleaching (FRAP) experiments with Utr–GFP show extensive ($83 \pm 22\%$, $N = 5$) turnover within <3 s (Supplementary Fig. 6c, d), suggesting that the actomyosin flow is mainly determined by Myo-II contractility on a fast-recycling, 'permissive' actin substrate.

We then addressed whether medial pulses are transferred to the junctional cortex and cause the formation of junctional pulses. As medial MRLC–GFP is slightly (500–1,000 nm) more apical than junctional MRLC–GFP, confocal sections distinguished the two pools and showed fusion of medial Myo-II (red) to the cortex and formation of a junctional pulse (green) (Fig. 3d left, right; Supplementary Movie 7). No transfer of medial pulses occurred to the adjacent junction following their ablation (Fig. 2c right). Moreover, photobleaching of MRLC–GFP along a junction (Supplementary Fig. 7a, b; Supplementary Movie 8) indicates two sources of exchange: pre-existing Myo-II patches are rapidly and strongly recovered ($72 \pm 6\%$), consistent with previous reports[7]; new junctional patches form *de novo* where medial Myo-II clusters fuse with junctions.

Junctional Myo-II pulses are delayed by ~8 s relative to medial ones (Supplementary Fig. 3), reflecting a speed of transfer of ~0.125 μm s^{-1}, which is similar to the direct flow speed measurements ($0.11 \pm 0.03 \, \mu m \, s^{-1}$).

Thus, medial and junctional actomyosin networks have tightly coordinated and hierarchically organized mechanical functions. Medial pulses flow to and produce steps of shrinkage of the adjacent 'vertical' junctions. They subsequently fuse with junctions and sustain junctional Myo-II accumulation, which stabilizes junction length. This flow and transfer are planar polarized, and drive junctional planar polarity and cell intercalation.

What controls the planar polarized flow of medial Myo-II pulses to vertical junctions? Mechanical anchoring of actomyosin networks at AJs is essential for force production during cell morphogenesis[6,17–19,25]. The medial network is also potentially connected to the apical plasma membrane given its tight apposition (Supplementary Fig. 8). Imaging of the apical plasma membrane with palmitoylated YFP (GAP43–Venus) revealed however a flat apical surface in the medial part of intercalating cells with few, small protrusions (Supplementary Fig. 9, Supplementary Movie 9), unlike apically constricting mesoderm cells where the plasma membrane is strongly ruffled (Supplementary Movie 10). These protrusions display local jitter but no aggregation or flow patterns characteristic of the underlying actomyosin network, suggesting moderate coupling (Supplementary Fig. 9, Supplementary Movie 9). Co-imaging of GAP43–Cherry and Utr–GFP shows that small protrusions and F-actin had un-correlated trajectories (Supplementary Movie 11) or moved at different speeds (Supplementary Movie 12; 3.7-fold reduced lateral dynamics ($0.03 \pm 0.015 \, \mu m \, s^{-1}$, Supplementary Fig. 9) compared to MRLC–GFP or Utr–GFP ($0.11 \pm 0.03 \, \mu m \, s^{-1}$, Fig. 3c)). Therefore, the apical surface and the medial actomyosin network are in contact but moderately coupled.

This suggested that the anisotropic actomyosin flow may largely depend on the distribution of junctional anchoring points. This requires E-cadherin/β-Catenin complexes at AJs and depends on α-Catenin[18,26]. E-cadherin/β-Catenin/α-Catenin complexes are planar polarized[5] (not shown), such that medial pulses flow towards regions with lower amounts of E-cadherin complexes. The level of E-cadherin along 'vertical' relative to adjacent junctions (E-cadherin anisotropy, Fig. 4a left) is also fluctuating (Fig. 4a middle). Moreover, the onset of medial pulses coincided with the time when E-cadherin anisotropy reached a local maximum (Fig. 4a middle, right) raising the possibility that E-cadherin anisotropy may orient the actomyosin flow. Reduction of E-cadherin by RNAi causes the disappearance of medial Myo-II (Fig. 4b top, c top; Supplementary Movies 13, 14). The junctional Myo-II level is consequently strongly reduced and no longer planar polarized (Fig. 4b bottom, c bottom). We reasoned that reducing the levels of α-Catenin by RNAi should attenuate coupling more subtly. α-Catenin RNAi reduces the number of E-cadherin clusters at AJs and disrupts interactions with junctional F-actin[18]. Moreover, the distribution of E-cadherin is no longer planar polarized in α-CateninRNAi embryos (Fig. 4e, Supplementary Fig. 10). This is associated with a loss of medial (Fig. 4f, Supplementary Movie 15) and junctional (Fig. 4d

top, bottom) Myo-II planar polarity. Thus, the planar polarized distribution of E-cadherin/β-Catenin/α-Catenin complexes biases the flow of medial Myo-II and junctional polarization.

In addition to Myo-II contractility, flow requires (1) crosslinkers between filaments to transmit tension within the medial meshwork, and (2) coupling at the cortex to E-cadherin/β-Catenin/α-Catenin complexes. Increased levels of E-cadherin in 'transverse' junctions may change properties of the actin network (for example, crosslinking/viscosity) and inhibit internal transmission of contractile forces and hence prevent D–V oriented flow. To test this, we disrupted the force balance within the medial actomyosin network by focal ablation (Fig. 4g top, bottom), and imaged the redistribution of medial clusters. If increased E-cadherin levels at transverse junctions inhibit tension transmission along the D–V axis, then medial pulses should not flow in this direction following ablation. However, we observed that Myo-II medial clusters flowed radially and away from the point of ablation towards the junctions (velocity $v = 0.05 \pm 0.01\,\mu\text{m s}^{-1}$) in 100% of cases ($N = 25$), even towards transverse junctions (12/25 cases, Fig. 4g top, bottom; Supplementary Fig. 11; Supplementary Movie 16). Focal ablation of the actin meshwork produces a local hole, which expands radially (Supplementary Movie 17). This argues that transverse
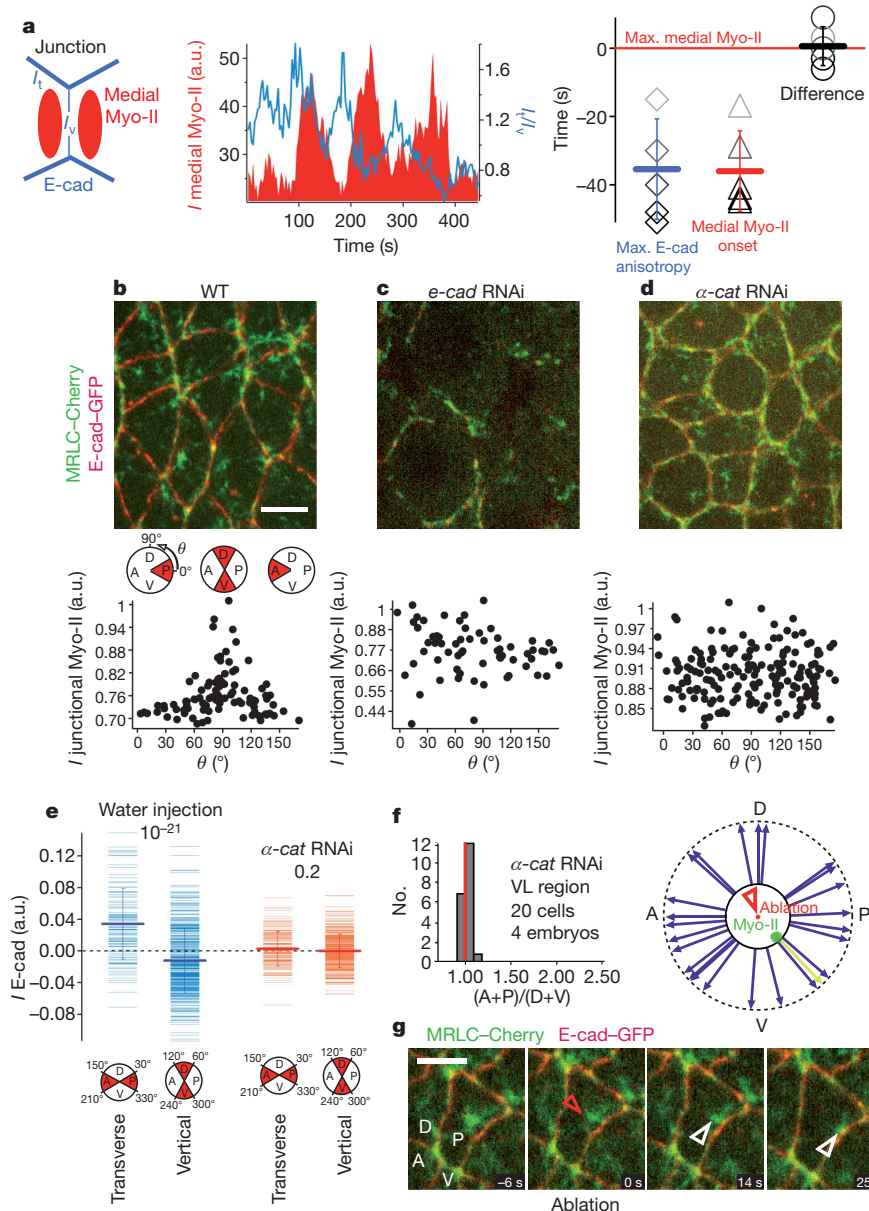
Figure 4 | E-cadherin planar polarity orients medial Myosin-II flow. a, Left and middle: medial MRLC–Cherry average intensity (red) and E-cad–GFP polarity (blue) as a function of time. E-cad–GFP polarity is the ratio of its mean intensity in transverse (I_t) and vertical (I_v) junctions. Right: chronology of events taking as a reference medial Myo-II intensity maximum. Delays between events are obtained by correlation; shown are mean and s.d. The difference is in black. **b–d**, Top row: Myo-II and E-cad in control (**b**), *e-cad* RNAi (**c**) and *α-cat* RNAi (**d**) embryos. Bottom row: average intensity of junctional Myo-II as a function of the angle (θ) of the junctions with respect to the A/P axis. **e**, Top: Comparison between normalized E-cad–GFP average intensity ($= (I_j - \bar{I})/\bar{I}$) of transverse versus vertical junctions for water injected (blue) and *α-cat* RNAi embryos (orange); *P* values are shown (Student's T test). I_j, mean intensity at a junction; $\bar{I}$, mean intensity of all junctions in a cell. Diagram at bottom indicates the angles of vertical and transverse junctions with respect to the A–P axis. **f**, Histogram of average medial Myo-II intensity as in Fig. 2a, right, for *α-cat* RNAi embryos. WT/*α-cat* RNAi: *P* = 0.0006 (Student's T test). **g**, Bottom: movement of a Myo-II cluster (white arrowhead) following nearby focal ablation (red arrowhead). Top right: diagram showing the centrifugal directions of the trajectories followed by Myo-II clusters ($N = 25$). Scale bars, 5 μm.

junctions do not inhibit flow *per se* and that flow directionality emerges from the properties of the actomyosin meshwork integrated over the entire apical surface.

The mechanical properties of the medial actomyosin network are locally defined by Myo-II contractility (concentration, affinity, duty cycle), tension transmission within the network (crosslinking), and viscous resistance to deformations (interactions between filaments)[27,28]. Moreover, these properties fluctuate owing to protein turnover and interactions. E-cadherin is known to anchor[18,26] and modify actin dynamics[29,30]. Our results suggest that the polarized distribution of E-cadherin may control the actomyosin flow pattern by spatially modulating mechanical properties of the actin network.

Current models of epithelial morphogenesis centre on Myo-II steady state distribution and associated contractile forces[1,2,4,5,7,10,12,14,15]. Our data show however that cell deformations cannot be simply derived from the Myo-II distribution itself, but from two central features of actomyosin dynamics, namely concentration (pulses) and movement (flow). Pulsed dynamics defines the rhythm and possibly the speed of deformation. Flow pattern, which in the case of intercalation is anisotropic, dictates the orientation of cell deformation (Supplementary Fig. 12). Flows of Myo-II foci have been reported in the one-cell stage *C. elegans* embryo[22,24], pointing to a more general property of actomyosin networks[15,23]. An important future avenue of research will be to investigate what properties of actin networks control Myo-II flow dynamics in different systems.

METHODS SUMMARY

Mutants and constructs. To visualize Myosin-II we used MRLC fused to eGFP or mCherry and rescuing a protein null *sqh*[AX3] mutant. The following stocks were used: *sqh*[AX3]; *sqh-MRLC::GFP (II)* and *sqh*[AX3]; *ubi-e-cad::GFP, sqh-MRLC::mCherry*. The plasmid coding for the fusion of eGFP and the actin binding domain of human Utrophin was obtained from W. Bement. The Utr–GFP clone was cloned by PCR in a pUASp destination vector (Fig. 1, Supplementary Fig. 1, Supplementary Movies 1a, b) or under the *sqh* promoter (Fig. 3, Supplementary Fig. 2, Supplementary Movie 3). The construct were verified by sequencing. To label the plasma membrane, we used a fusion between the palmitoylated GAP43 protein and YFP/Venus expressed by the GAL4/UAS system with the maternal tubGAL4VP16 driver line. GAP43–Cherry was expressed under the *sqh* promoter.

RNA interference. We generated by PCR dsRNA probes directed against *Krüppel*, α-catenin and *e-cadherin* as described in refs 4, 18.

Time-lapse imaging. Embryos were prepared and imaged using a spinning disc confocal system (Perkin Elmer) on an inverted Nikon microscope with 100× oil immersion objective. Nano-ablation was performed using a home-built set-up[12]. Fluorescence recovery after photobleaching (FRAP) measurements were performed as in Supplementary Fig. 6 using a confocal LSM510 (Zeiss) with a Plan-Apochromat 100× oil objective and an argon laser (488 nm).

Image analysis and quantifications. Intensity measurements, cross-correlation analysis, time-delays analysis and PIV analysis are detailed in Online Methods and in Supplementary Figs 3, 4.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lecuit, T. & Lenne, P. F. Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis. *Nature Rev. Mol. Cell Biol.* **8,** 633–644 (2007).
2. Montell, D. J. Morphogenetic cell movements: diversity from modular mechanical properties. *Science* **322,** 1502–1505 (2008).
3. Hutson, M. S. *et al.* Forces for morphogenesis investigated with laser microsurgery and quantitative modeling. *Science* **300,** 145–149 (2003).
4. Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429,** 667–671 (2004).
5. Blankenship, J. T., Backovic, S. T., Sanny, J. S., Weitz, O. & Zallen, J. A. Multicellular rosette formation links planar cell polarity to tissue morphogenesis. *Dev. Cell* **11,** 459–470 (2006).
6. Dawes-Hoang, R. E. *et al.* Folded gastrulation, cell shape change and the control of myosin localization. *Development* **132,** 4165–4178 (2005).
7. Fernandez-Gonzalez, R., Simoes Sde, M., Roper, J. C., Eaton, S. & Zallen, J. A. Myosin II dynamics are regulated by tension in intercalating cells. *Dev. Cell* **17,** 736–743 (2009).
8. Hildebrand, J. D. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. *J. Cell Sci.* **118,** 5191–5203 (2005).
9. Krieg, M. *et al.* Tensile forces govern germ-layer organization in zebrafish. *Nature Cell Biol.* **10,** 429–436 (2008).
10. Monier, B., Pelissier-Monier, A., Brand, A. H. & Sanson, B. An actomyosin-based barrier inhibits cell mixing at compartmental boundaries in *Drosophila* embryos. *Nature Cell Biol.* **12,** 60–65 (2010).
11. Nishimura, T. & Takeichi, M. Shroom3-mediated recruitment of Rho kinases to the apical cell junctions regulates epithelial and neuroepithelial planar remodeling. *Development* **135,** 1493–1502 (2008).
12. Rauzi, M., Verant, P., Lecuit, T. & Lenne, P. F. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nature Cell Biol.* **10,** 1401–1410 (2008).
13. Skoglund, P., Rolo, A., Chen, X., Gumbiner, B. M. & Keller, R. Convergence and extension at gastrulation require a myosin IIB-dependent cortical actin network. *Development* **135,** 2435–2444 (2008).
14. Landsberg, K. P. *et al.* Increased cell bond tension governs cell sorting at the *Drosophila* anteroposterior compartment boundary. *Curr. Biol.* **19,** 1950–1955 (2009).
15. Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457,** 495–499 (2009).
16. Nishimura, T. & Takeichi, M. Remodeling of the adherens junctions during morphogenesis. *Curr. Top. Dev. Biol.* **89,** 33–54 (2009).
17. Gorfinkiel, N. & Arias, A. M. Requirements for adherens junction components in the interaction between epithelial tissues during dorsal closure in *Drosophila. J. Cell Sci.* **120,** 3289–3298 (2007).
18. Cavey, M., Rauzi, M., Lenne, P. F. & Lecuit, T. A two-tiered mechanism for stabilization and immobilization of E-cadherin. *Nature* **453,** 751–756 (2008).
19. Sawyer, J. K., Harris, N. J., Slep, K. C., Gaul, U. & Peifer, M. The *Drosophila* afadin homologue Canoe regulates linkage of the actin cytoskeleton to adherens junctions during apical constriction. *J. Cell Biol.* **186,** 57–73 (2009).
20. Irvine, K. D. & Wieschaus, E. Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development* **120,** 827–841 (1994).
21. Zallen, J. A. & Wieschaus, E. Patterned gene expression directs bipolar planar polarity in *Drosophila. Dev. Cell* **6,** 343–355 (2004).
22. Hird, S. N. & White, J. G. Cortical and cytoplasmic flow polarity in early embryonic cells of *Caenorhabditis elegans. J. Cell Biol.* **121,** 1343–1355 (1993).
23. Bertet, C., Rauzi, M. & Lecuit, T. Repression of Wasp by JAK/STAT signalling inhibits medial actomyosin network assembly and apical cell constriction in intercalating epithelial cells. *Development* **136,** 4199–4212 (2009).
24. Munro, E., Nance, J. & Priess, J. R. Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior-posterior polarity in the early *C. elegans* embryo. *Dev. Cell* **7,** 413–424 (2004).
25. Martin, A. C., Gelbart, M., Fernandez-Gonzalez, R., Kaschube, M. & Wieschaus, E. F. Integration of contractile forces during tissue invagination. *J. Cell Biol.* **188,** 735–749 (2010).
26. Yonemura, S., Wada, Y., Watanabe, T., Nagafuchi, A. & Shibata, M. α-Catenin as a tension transducer that induces adherens junction development. *Nature Cell Biol.* **12,** 533–542 (2010).
27. Bendix, P. M. *et al.* A quantitative analysis of contractility in active cytoskeletal protein networks. *Biophys. J.* **94,** 3126–3136 (2008).
28. Koenderink, G. H. *et al.* An active biopolymer network controlled by molecular motors. *Proc. Natl Acad. Sci. USA* **106,** 15192–15197 (2009).
29. Kobielak, A., Pasolli, H. A. & Fuchs, E. Mammalian formin-1 participates in adherens junctions and polymerization of linear actin cables. *Nature Cell Biol.* **6,** 21–30 (2004).
30. Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. & Weis, W. I. α-Catenin is a molecular switch that binds E-cadherin-β-catenin and regulates actin-filament assembly. *Cell* **123,** 903–915 (2005).

Author Contributions M.R. made the original observations of pulsed contractility and flow. M.R. and T.L. planned the project, and analysed the data together with P.-F.L; M.R. conducted the experiments except for FRAP experiments on MRLC–GFP, which were performed by P.-F.L and T.L; P.-F.L and M.R. developed the nano-ablation system. T.L. and P.-F.L. wrote the manuscript together with M.R. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.L. (lecuit@ibdml.univ-mrs.fr) or P.-F.-L. (lenne@ibdml.univ-mrs.fr).

METHODS

Fly stocks and constructs. *Drosophila* MRLC is encoded by *spaghetti-squash (sqh)*. All experiments visualizing dynamics of MRLC were looking at MRLC fused to either eGFP or mCherry under the *sqh* promoter and rescuing a protein null *sqh*^{AX3} mutant[31]. E-cad–GFP was expressed under the ubiquitin promoter *ubi-Ecad::GFP* and rescues a null *e-cad/shotgun* mutant[32].

The following fly stocks were used. Figure 1, Supplementary Fig. 1 and Supplementary Movies 1a, b: matGAL4(67) UASp-Utr–GFP (recombinant on II). *sqh*^{AX3}; *sqh-MRLC::GFP (II)* (generous gift of R. Karess) and *sqh*^{AX3}; *ubi-E-cad::GFP, sqh-sqh::mCherry* (recombinant on II). Figure 3, Supplementary Fig. 2, Supplementary Movie 3: *sqh*^{AX3}; *sqh-MRLC::mCherry, sqh-Utr::GFP* (recombinant on II). *sqh-MRLC::mCherry* (on II) is a gift from A. Martin and E. Wieschaus.

The plasmid coding for the fusion of eGFP and the actin binding domain of human Utrophin was obtained from W. Bement[33]. The Utr–GFP fusion was PCR amplified and inserted in the p221DONR GATEWAY plasmid (Invitrogen). The fusion was recombined in a pUASp GATEWAY destination vector (pPW, from T. Murphy, Carnegie Institute) for expression under the maternal tubGAL4VP16 driver line (67Gal4) in Fig. 1, Supplementary Fig. 1 and Supplementary Movies 1a, b, or expression under the *sqh* promoter in Fig. 3, Supplementary Fig. 2 and Supplementary Movie 3. In the latter case, Utr–GFP expression is lower than under the Gal4 system, hence only reveals brighter structures (puncta) also visible in Supplementary Movie 1a and b and Fig. 1d and not individual filaments contrary to Fig. 1d.

To label the plasma membrane we used a fusion between the palmitoylated GAP43 protein and the YFP variant Venus[34] expressed by the GAL4 UAS system with the maternal tubGAL4VP16 driver line. GAP43–Cherry was constructed similarly and expressed under the *sqh* promoter as in ref. 31.

RNAi interference. We generated by PCR dsRNA probes directed against *Krüppel*, α-catenin, and *e-cadherin* using the following primers. The underlined sequence is the T7 promoter. The sequence not underlined corresponds to the template sequence. e-cadherin: 533 nucleotides, between +1475 to +2008 from ATG. E-cad-T7-F, <u>taatacgactcactataggg</u>agaccacgagtctctttgataatggcgagc. E-cad-T7-R, <u>taatacgactcactatagggagacca</u>cggtttccatcgttctggtgaatc. α-catenin: 728 nucleotides, between +81 to +808 from ATG. α-Cat-T7-F, <u>taatacgactcactataggg</u>cacaatgtcagttgaaaaaacacttg. α-Cat-T7-R, <u>taatacgactcactatagggg</u>gttgggatgactttccttgggc aac. Krüppel: 775 nucleotides, between +491 to +1266 from ATG. Kr-T7-F, <u>taatacgactcactataggg</u>gagaccacggagtttcagaccgagatcagca. Kr-T7-R, <u>taatacgactcactat agggagaccacaga</u>gagctggctccatcttcagaca. Embryos were injected as described in ref. 35.

Time lapse imaging. Embryos were prepared and imaged as detailed in ref. 36, using a spinning disc confocal system (Perkin Elmer) on an inverted Nikon microscope with 100×/1.4 oil immersion objective.

Nano-ablation experiments. We performed nano-dissection experiments with a home-built system. A near-infrared (NIR, 1,030 nm) femtosecond (fs) laser at 50 MHz repetition rate (t-Pulse, Amplitude Systems) was coupled to an inverted microscope (Eclipse TE 2000-E, Nikon). A fast multicolour confocal imaging system, based on the Yokogawa spinning disk (Ultraview ERS, Perkin Elmer), was also mounted at a side port of the microscope. Local ablation and fast fluorescence imaging were thus possible simultaneously. The NIR-fs laser beam is expanded through a ×5 telescope and is aligned with the microscope optical path with a dichroic mirror (FF01-750/SP, Semrock) immediately below the objective lens (×60/1.2, water immersion, Plan Apo VC, Nikon). The collimated beam fills the back aperture of the objective lens which transmits 68% of the incoming NIR light. Nano-dissections of medial Myo-II were performed by exposing this structure to the tightly focused laser during 1–3 ms with an average power of 360 mW at the back aperture of the objective. Exposure time was controlled by an automated 1.5-mm-diameter mechanical shutter (LS2, Uniblitz). The sample was positioned over the tightly focused laser beam thanks to a computer-controlled mechanical stage (Scan IM with a Tango2-Desktop controller, Marzhaüser). A very similar set-up has already been shown to allow sub-cellular ablations[12].

Fluorescence intensity measurements. The intensity of the medial Myo-II is defined as the sum of average intensities of two regions of interest (ROIs) close to the junction (the centre of the elliptical ROIs were ~1 μm away from the junction, Fig. 2a in red). The intensity of the junctional Myo-II is defined as the average intensity of a 500-nm-wide stripe along the junction (Fig. 2a in green). The E-cad anisotropy is the average intensity of a 500-nm-wide stripe along transverse junctions divided by the average intensity measured along the vertical junction (Fig. 4a in blue). Intensity measurements were made by using ImageJ (1.39p version). Analysis were done on time lapse movies (one frame every 1–3 s). For

each frame, 6–10 z-planes were imaged over 3 μm. For long time lapse imaging (>200 s), bleach correction was performed by using ImageJ.

Cross-correlation analysis. Cross-correlation was performed applying Igor Pro (Wavemetrics) cross-correlation function. This function is given by:

$$C(\tau) = \int_0^T f(t)g(t+\tau)\mathrm{d}t$$

where T represents the overall time over which measurements were made, $f(t)$ and $g(t)$ the two cross-correlated functions (taking f as reference), and τ the time delay.

The basal signal $f_{\min}$ and $g_{\min}$ were subtracted from f and g functions respectively before cross-correlation. The final cross-correlation function was normalized as follows:

$$C_N(\tau) = \frac{\int_0^T (f(t)-f_{\min})(g(t+\tau)-g_{\min})\mathrm{d}t}{\sqrt{\int_0^T (f(t)-f_{\min})^2\mathrm{d}t}\cdot\sqrt{\int_0^T (g(t)-g_{\min})^2\mathrm{d}t}}$$

Time delay measurements. In Fig. 2b all time delays were measured by cross-correlation. Cross-correlation analysis was assessed by performing a measure of delays between peaks for each cluster of events (for example, Fig. 2 shows three clusters of events) (see Supplementary Figs 1 and 2 top middle panel). When correlating with contraction rate functions, curves were smoothed by using a binomial algorithm implemented in Igor Pro software. For this analysis five cases of fully intercalating cells (corresponding to 15 clusters of events) were taken from five different wild-type MRLC–GFP embryos. Time lapse movies were taken at a rate of 1 frame s^{-1}. Each frame consisted of a z-stack of 3 μm (images spaced by 500 nm). Time lapse ranged between 200 and 500 s. Time delays in Fig. 4a, right, were determined as follows. The delay between the E-cad anisotropy peak and the medial Myo-II intensity peak was measured by cross-correlation (the medial Myo-II intensity curve was taken as reference). The time onset of medial Myo-II intensity pulses with respect to medial Myo-II intensity peak maxima was determined from the autocorrelation of the medial Myo-II intensity, which provides a measure of the average pulse duration, and therefore a measure of the average delay between pulse onset and pulse intensity peak. Auto-correlation analysis was assessed by performing a measure of delays for each cluster of events (see Supplementary Fig. 2 bottom) as for cross-correlation analysis. For this analysis five cases of intercalating cells were taken from five different wild-type MRLC–Cherry / E-cad–GFP embryos. Time lapse movies (one frame every 3 s) of both MRLC–Cherry and E-cad–GFP were taken. Each frame consisted of a z-stack of 3 μm (images spaced by 500 nm). Time lapse ranged between 200 and 500 s. Igor Pro software was used for all time delay measurements.

PIV analysis. PIV was determined with the Mathlab toolbox (procedure MatPIV) developed by J. K. Sveen.

Fluorescence recovery after photobleaching. Fluorescence recovery after photobleaching (FRAP) measurements were performed as in Supplementary Fig. 6 using a confocal LSM510 (Zeiss) with a Plan-Apochromat 100×/1.3 oil objective and an argon laser (488 nm). Before and after photobleaching, images were acquired at low laser power (0.1% AOTF, Acousto Optic Tunable Filter) to avoid bleaching and with a pixel size of 40 nm. Photobleaching was performed for 0.9 s at full laser power over an ROI with 1 μm diameter. Fluorescence recovery was then recorded for 50 s. In Supplementary Fig. 7, we used the photokinesis unit of a Perkin Elmer confocal system for FRAP and the region of interest is a line 5 μm long.

31. Royou, A., Field, C., Sisson, J. C., Sullivan, W. & Karess, R. Reassessing the role and dynamics of nonmuscle myosin II during furrow formation in early *Drosophila* embryos. *Mol. Biol. Cell* **15**, 838–850 (2004).
32. Oda, H. & Tsukita, S. Real-time imaging of cell-cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114**, 493–501 (2001).
33. Burkel, B. M., von Dassow, G. & Bement, W. M. Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. *Cell Motil. Cytoskeleton* **64**, 822–832 (2007).
34. Mavrakis, M., Rikhy, R. & Lippincott-Schwartz, J. Plasma membrane polarity and compartmentalization are established before cellularization in the fly embryo. *Dev. Cell* **16**, 93–104 (2009).
35. Pilot, F., Philippe, J. M., Lemmers, C. & Lecuit, T. Spatial control of actin organization at adherens junctions by a synaptotagmin like protein Btsz. *Nature* **442**, 580–584 (2006).
36. Cavey, M. & Lecuit, T. Imaging cellular and molecular dynamics in live embryos using fluorescent proteins. *Methods Mol. Biol.* **420**, 219–238 (2008).

Light–avoidance–mediating photoreceptors tile the *Drosophila* larval body wall

Yang Xiang[1], Quan Yuan[1], Nina Vogt[2], Loren L. Looger[3], Lily Yeh Jan[1] & Yuh Nung Jan[1]

Photoreceptors for visual perception, phototaxis or light avoidance are typically clustered in eyes or related structures such as the Bolwig organ of *Drosophila* larvae. Unexpectedly, we found that the class IV dendritic arborization neurons of *Drosophila melanogaster* larvae respond to ultraviolet, violet and blue light, and are major mediators of light avoidance, particularly at high intensities. These class IV dendritic arborization neurons, which are present in every body segment, have dendrites tiling the larval body wall nearly completely without redundancy. Dendritic illumination activates class IV dendritic arborization neurons. These novel photoreceptors use phototransduction machinery distinct from other photoreceptors in *Drosophila* and enable larvae to sense light exposure over their entire bodies and move out of danger.

Light sensing is critical for animal life. Whereas image-forming visual perception allows animals to identify and track mates, predators and prey, non-image-forming functions regulate pupil reflex, phototaxis and circadian entrainment[1,2]. In addition to eyes[1,2], extra-ocular photoreceptors exist[1–5]. For example, many eyeless or blinded animals can sense illumination of their body surfaces[3–5]. Birds possess deep-brain photoreceptors in their hypothalamus[6], and extra-ocular photoreceptors are required for magnetic orientation of amphibians[7]. Recent studies demonstrate that eyeless animals such as *Caenorhabditis elegans* nonetheless have photoreceptors controlling light avoidance[8–10].

Drosophila larvae spend most of the time feeding by digging into food. Light avoidance is a crucial behaviour to minimize body exposure. When tested in groups in a dark/light choice assay, *Drosophila* larvae prefer darkness[11,12]. This behaviour requires the pair of Bolwig organs on the larval head[12]; that is, primitive eye structures each comprised of 12 photoreceptors expressing Rh5 or Rh6, rhodopsins sensing blue and green light, respectively[13].

Cells besides Bolwig organs contribute to photoavoidance

We designed a photoavoidance assay for a single larva with sunlight-level intensities ($\sim$1 mW mm^{-2} in San Francisco on a clear day in June, consistent with previous reports[14]). Wild-type *Drosophila* larvae showed avoidance of a white light spot of 0.57 mW mm^{-2} (Fig. 1a and Supplementary Movie 1). Surprisingly, similar avoidance (Fig. 1b and Supplementary Movie 2) was exhibited by larvae with their Bolwig organs ablated by the pro-apoptotic gene *Head involution defective* (*Hid*; also called *Wrinkled* (*W*))[15] expressed via the Bolwig-organ-specific promoter *Glass Multimer Reporter* (*GMR*)[16] (Supplementary Fig. 1a). Lower light intensities elicited less photoavoidance of wild-type animals, and even less of Bolwig-organ-ablated animals (Fig. 1c). However, at light intensities of 0.57 mW mm^{-2} or higher, *GMR-Hid* larvae showed avoidance comparable to wild-type animals ($P > 0.05$). Thus, although the Bolwig organs are responsible for dim light avoidance and dark congregation[12], *Drosophila* larvae must contain extra-ocular photoreceptors.

Testing the wavelength dependence of photoavoidance using band-pass filters letting through ultraviolet ($\sim$360 nm; Fig. 1d), violet

($\sim$402 nm; Fig. 1e), blue ($\sim$470 nm; Fig. 1f), green ($\sim$525 nm; Fig. 1g) or red light ($\sim$620 nm; Fig. 1h), we found that wild-type animals showed increased photoavoidance with higher light intensity (Fig. 1d–h), and were most sensitive to blue, violet and ultraviolet, and largely unresponsive to green and red light. Bolwig-organ-ablated animals showed less photoavoidance at low light intensity, but exhibited nearly normal avoidance response to high-intensity, short-wavelength light (Fig. 1d–f), demonstrating the existence of light-sensitive cells in addition to Bolwig organs. Because there was no detectable temperature increase associated with 0.11 mW mm^{-2} violet light (Supplementary Fig. 2)—which triggered avoidance in nearly 80% of the wild-type and *GMR-Hid* animals—and animals showed little response to high-intensity green or red light but strongly avoided low-intensity short-wavelength light (Fig. 1 d–h), light avoidance probably involves wavelength-dependent photoreceptors but not local heating.

Class IV neurons tiling larval body wall sense light

Given the report of diffusely distributed dermal photoreceptors triggering shadow reaction[3–5], we tested whether sensory neurons in the larval body wall could be candidate photoreceptors. Using GCaMP3, a genetically encoded calcium indicator[17–19], we found that blue light delivered for 5 s to the dorsal cluster (Fig. 2a, see Fig. 3c for whole larva image) generated a marked fluorescence increase specifically in the soma, axon and dendrites of ddaC, a class IV dendritic arborization neuron (Fig. 2b, e, f), but not in nearby sensory neurons (Fig. 2b, e, f). ddaC also responded to ultraviolet light (which also caused photo-bleaching), but not to green light (Fig. 2c–f). There were also Ca^{2+} increases specifically in class IV dendritic arborization neurons of the ventral and lateral cluster (V'ada and VdaB, respectively) in response to ultraviolet and blue light, but not green light (Supplementary Figs 3 and 4). Similar GCaMP3 fluorescence responses were seen in class IV dendritic arborization neurons in body segments from head to tail (Supplementary Fig. 5).

Extracellular recording further revealed a progressive increase in action potential frequency when the dorsal class IV dendritic arborization neuron ddaC was illuminated with increasing intensity of blue light (Fig. 2g), ultraviolet light and violet light, but not red light
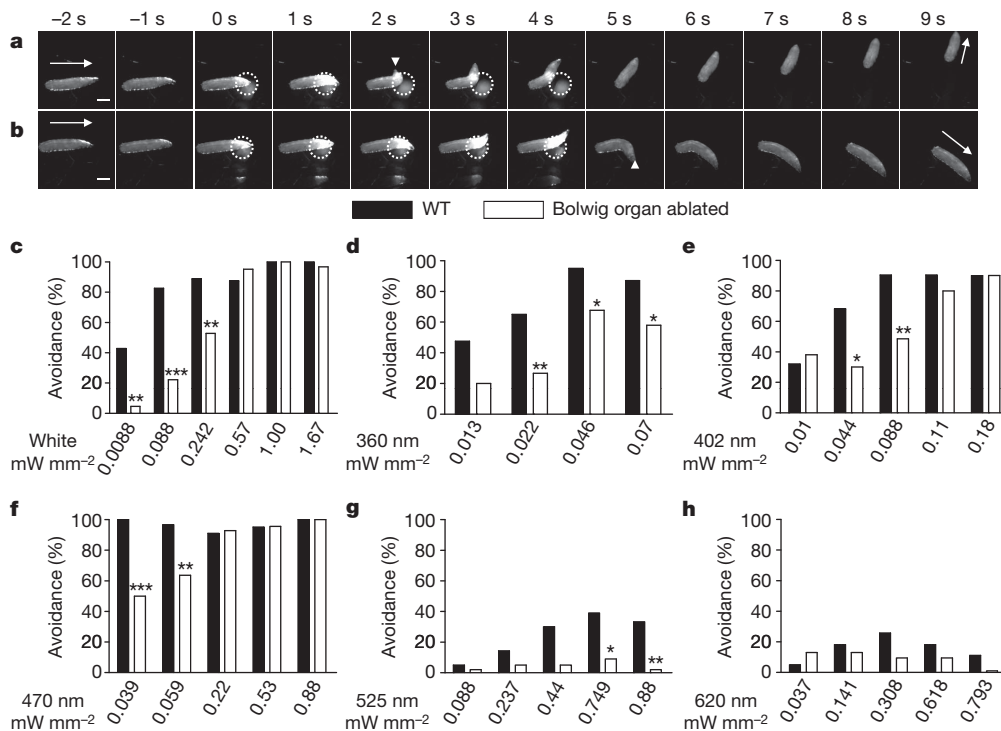
Figure 1 | Photoreceptors in addition to Bolwig organs contribute to photoavoidance. **a, b**, Examples of light avoidance of wild-type (**a**) and *GMR-Hid* (**b**) larvae exposed to white light (0.57 mW mm^{-2}) applied from 0 to 5 s. The light spot is indicated by the dotted circle. The arrow indicates the direction of larval locomotion; arrowheads at 2 s (**a**) and 5 s (**b**) indicate larval head turning. **c–h**, Percentage of animals avoiding white light (**c**), light of 360 nm (ultraviolet; **d**), 402 nm (violet; **e**), 470 nm (blue; **f**), 525 nm (green; **g**) and 620 nm (red; **h**) at different intensities. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, two-tailed Fisher exact test. Twenty to forty larvae were tested for each condition. Scale bar: 1 mm (**a, b**), shown at −2 s.

(Supplementary Fig. 6). Responses were: 340 nm > 380 nm > 402 nm > 470 nm ≫ 525 nm or 620 nm light (Fig. 2h).

The wavelength dependence of ddaC firing rate increase was similar to that observed with GCaMP3 imaging and the light avoidance behavioural assay. The latency between the onset of light stimulation and action potential burst firing decreased with higher light intensity, and was as short as 1 s with bright illumination (Supplementary Fig. 6). When illuminated with 1.4 mW mm^{-2} of white light (approximating sunlight), ddaC neurons in the dorsal cluster showed a significant firing increase (Fig. 2i). Similar robust activation of ventral (VdaB) and lateral (V′ada) class IV dendritic arborization neurons was induced by 52.8 mW mm^{-2} blue light (Supplementary Fig. 7a, b). The response of class IV dendritic arborization neurons was similar regardless of their location along the body axis (data not shown), as in the case of GCaMP3 imaging (Supplementary Fig. 5).

We did not observe any significant effects of light on firing rate of class I or III dendritic arborization neurons (Supplementary Fig. 7c, d, $P > 0.05$). Because class I dendritic arborization neurons progressively increased their firing rate as the temperature was raised above 30 °C, whereas class IV dendritic arborization neurons showed an abrupt increase of firing rate only above 40 °C (Supplementary Fig. 8), thermal responses cannot account for the light-induced increase of firing in class IV but not class I dendritic arborization neurons. Moreover, application of 10 μM H_2O_2, which elevates the reactive oxygen species (ROS) level in *Drosophila* larvae[20], had no effect on the firing rate of class IV dendritic arborization neurons (Supplementary Fig. 9). These studies demonstrate that ultraviolet, violet and blue light activate class IV dendritic arborization neurons in an intensity-dependent manner. Responses occur at sunlight-level intensities, are not induced by heat or ROS, correlate with behaviour, and are confined to this specific class of sensory neurons throughout the animal.

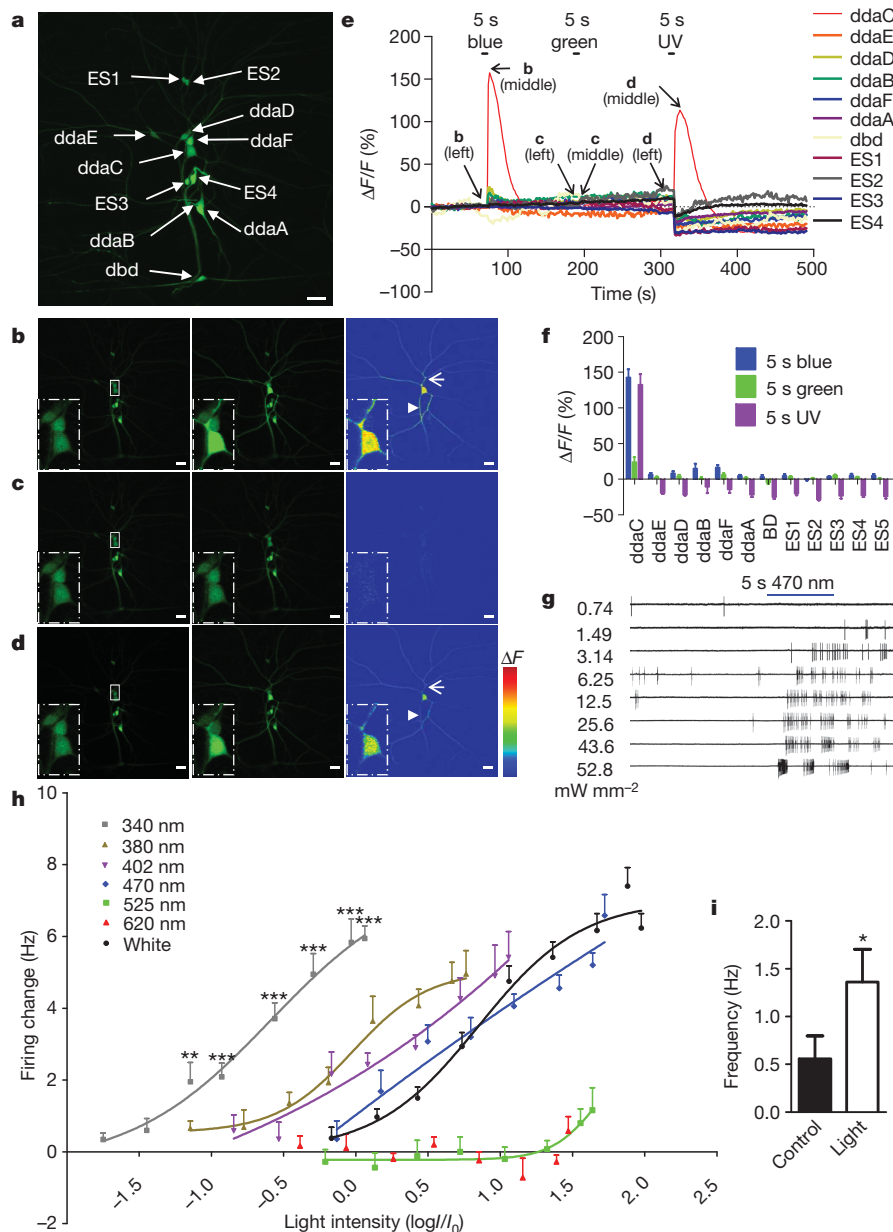## Light activates class IV neurons and dendrites in isolation

The dendritic arborization neurons have dendrites in contact with epithelial cells whereas their somas and axons are wrapped by glia[21]. To test whether class IV dendritic arborization neurons can sense light by themselves, we prepared primary neuronal cultures[22,23] from embryos expressing GCaMP3 and RFP specifically in class IV dendritic arborization neurons by means of *pickpocket-GAL4* (*ppk-GAL4*)[24].

Ultraviolet and blue light illumination of isolated class IV dendritic arborization neurons generated a robust increase of GCaMP3 signals (Fig. 3a and Supplementary Fig. 10). In contrast, cultured class III dendritic arborization neurons expressing GCaMP3 and RFP via *19-12-GAL4* yielded no light response (Fig. 3b). Thus, class IV dendritic arborization neurons have the intrinsic ability to detect light.

Dendrites of class IV dendritic arborization neurons tile the larval body wall with non-overlapping but complete coverage of the dendritic field[25,26] (Fig. 3c). Illumination of only the dendrites of class IV dendritic arborization neurons (Fig. 3d) with ultraviolet, violet and blue light, but not green or red light, activated the neurons (Fig. 3e). The activation spectrum is similar to that for illumination of the entire class IV dendritic arborization neurons (Fig. 2h), indicating the presence of phototransduction machinery in the dendrites.

Gr28b is critical for light transduction in class IV neurons

No defects in light response of class IV dendritic arborization neurons were found in available mutants of rhodopsins[13,27] and cryptochrome (*cry*)[28], as well as a mutant in *no receptor potential A* (*norpA*), which encodes phospholipase C (PLC), downstream of rhodopsins[29] (Fig. 4a). We then tested the *Drosophila* homologue of Lite-1, a *C. elegans* light sensor[8–10]. The closest homologue of *lite-1* in *Drosophila* is gustatory receptor 28b (*Gr28b*), annotated as encoding a gustatory G-protein-coupled receptor. Several *Gr28b-GAL4* lines carrying different promoter regions revealed consistent expression in all class IV dendritic arborization neurons, two sensory neurons in the lateral body wall, plus several neurons in the ventral nerve cord (Supplementary Fig. 11), as reported previously[30]. To test for the functional role of *Gr28b* in the light-induced electrophysiological responses, we recorded from class IV dendritic arborization neurons in Dmel\Mi{ET1}Gr28b[MB03888] (*MiET1*) and Dmel\PBac{PB}Gr28b[c01884] (*PBac*) larvae with P-element insertion into the *Gr28b* coding and intronic regions, respectively (http://flybase.org/reports/FBgn0045495.html). Whereas these P-element insertions did not alter the basal firing rate (data not shown), they caused a significant reduction in light-induced responses of class IV dendritic arborization neurons (Fig. 4b), as in hemizygous larvae carrying one *MiET1* allele and one deletion encompassing *Gr28b* (Df(2L)Exel7031; http://flybase.org/reports/FBab0037910.html) (Fig. 4c). The *MiET1* P-element inserts into the coding sequence

Figure 2 | Light activates class IV dendritic arborization neurons. a, Pre-stimulation image showing larval dorsal cluster sensory neurons (dbd, bipolar dendrite neuron; ddaD and ddaE, class I dendritic arborization neurons; ddaB, class II dendritic arborization neurons; ddaA and ddaF, class III dendritic arborization neurons; ddaC, class IV dendritic arborization neurons; ES, external sensory organ). Up is dorsal; right is anterior. For an atlas of the larval peripheral nervous system, see ref. 25. **b**, Responses of the dorsal cluster neurons in **a** to 5 s blue light (470 nm) illumination. The boxed area in the left panel and insets in all three panels show the somas of ddaC, ddaF and ddaD dendritic arborization neurons. Left, pre-stimulation; middle, post-stimulation; right, GCaMP3 intensity difference (middle panel minus left panel), with ddaC dendrites (arrow) and axon (arrowhead) marked. **c**, **d**, Similar experiments with 5 s green (546 nm; **c**) and ultraviolet light (365 nm; **d**) revealed ddaC activation by ultraviolet, but not green, light. Scale bar in **a**–**d**, 20 μm; colour scale in right panels of **b**–**d** shows dynamic range (0–4,095). **e**, Time course of somatic GCaMP3 signals of dorsal cluster neurons shown in **a**–**d**. Time frames are indicated. **f**, Summary of somatic fluorescence changes ($\Delta F/F$) of dorsal cluster neurons in response to 5 s light stimulation, $n = 7–16$. **g**, Example firing traces of ddaC in response to 5 s 470 nm blue light. **h**, Summary of firing frequency changes (average frequency of 5 s before light exposure subtracted from average frequency during 5 s of light exposure) of ddaC induced by white, 340, 380, 402, 470, 525 and 620 nm light. For clarity, significance is only shown for the 340 nm curve. Light intensity is reported as the log of (I normalized to $I_0 = 1$ mW mm^{-2}). Green (525 nm) or red (620 nm) light has no effect ($P > 0.05$). $n = 5–9$. **i**, Effect of 1.4 mW mm^{-2} white light on ddaC, average frequencies of 5 s before (control) and during the 5 s of light exposure (light) are plotted. $n = 6$. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; two-tailed paired t-test. All error bars indicate s.e.m.

common to all reported transcripts, and its mobilization for excision restored the light-induced response in class IV dendritic arborization neurons (Fig. 4d). Moreover, knockdown of Gr28b expression with *UAS-RNAi* driven by *ppk-GAL4* caused an overall reduction of light response of class IV dendritic arborization neurons (Fig. 4e). Taken together, our data indicate that Gr28b is expressed in class IV dendritic arborization neurons, and is required for proper light responses. Whether Gr28b is the direct photosensing molecule awaits further experimentation.

Sequence analysis revealed that Gr28b has a rhodopsin-like structure plus one extra transmembrane segment (Supplementary Fig. 12), raising the question of whether the Gr28b-dependent light response involves G-protein signalling. To test whether G-protein signalling is required in class IV dendritic arborization neurons, we applied the myristoylated βγ-binding peptide mSIRK, and found that the light response in class IV dendritic arborization neurons was significantly reduced (Supplementary Fig. 13). Thus, G-protein signalling is probably involved in the light response of class IV dendritic arborization neurons, similar to findings in *C. elegans*[10]. We further tested cyclic nucleotide-gated (CNG) channels, which are known to act downstream of Lite-1 and G proteins in *C. elegans*[10]. Unlike in *C. elegans*, blocking CNG

channels with L-*cis*-diltiazem in class IV dendritic arborization neurons had no effect on their light responses (Supplementary Fig. 14).

TrpA1 is required in light transduction in class IV neurons

Transient receptor potential (TRP) channels were first identified and characterized in the *Drosophila* compound eye[29,31], with TRP and TRP-like (*trpl*) having key roles in phototransduction[29]. However, our electrophysiological studies revealed no defects in the light response of class IV dendritic arborization neurons in *trpl* or *painless* mutant larvae (Supplementary Fig. 15).

TrpA1, a *Drosophila* homologue of mammalian TrpA, may function as a thermosensor in larvae and adults[32–34], and a receptor for reactive electrophiles such as allyl isothiocyanate (AITC)[35]. A *TrpA1* mutant exhibited normal basal firing in class IV dendritic arborization neurons (data not shown), but no light-induced firing increase (Fig. 4f). As reported previously[32], we detected strong TrpA1 immunoreactivity in several neurons in the larval brain but not in peripheral neurons (data not shown). We then performed MARCM (mosaic analysis with a repressible cell marker)[36], and found no light response in class IV dendritic arborization neurons lacking TrpA1 (Fig. 4g), and a significant reduction of light-induced firing of the heterozygous class
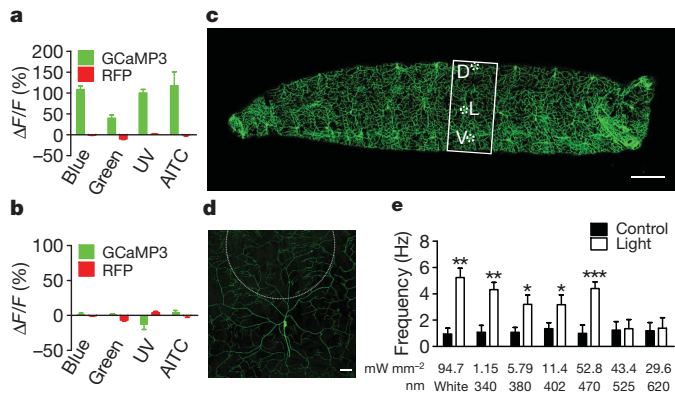
Figure 3 | Cell-autonomous activation of class IV dendritic arborization neurons by light. a, b, Quantification of somatic fluorescence changes ($\Delta F/F$) in response to 5 s light and 100 μM allyl isothiocyanate (AITC) stimulation of cultured class IV (**a**) and III (**b**) dendritic arborization neurons; RFP signals serve as control. $n = 10$–13 (light) and $n = 4$ (AITC) in **a**, $n = 9$ in **b**. **c,** Larva with class IV dendritic arborization neurons labelled with GFP by *ppk-GAL4*. Dendrites tile the body wall. Boxed area shows an abdominal hemi-segment; three dotted circles mark soma positions of D (dorsal, ddaC), L (lateral, V′ada) and V (ventral, VdaB) class IV dendritic arborization neurons, respectively. Up, dorsal; left, anterior. Scale bar, 200 μm. **d,** Illumination of dendrites within the dotted circle of GFP-labelled ddaC dendrites. Up, dorsal. Scale bar, 50 μm. **e,** Responses of ddaC with dendritic illumination. $n = 5$. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; two-tailed paired t-test. All error bars indicate s.e.m.

IV dendritic arborization neurons (Supplementary Fig. 16), indicating that TrpA1 is present in levels below immunodetection, but nonetheless of functional importance. In support of this notion, AITC caused strong activation of class IV dendritic arborization neurons, and this activation was abolished in the *TrpA1* mutant (Supplementary Fig. 17a). Moreover, *TrpA1* RNAi expression specifically in class IV dendritic arborization neurons eliminated the light-induced firing change (Supplementary Fig. 18). Taken together, our observations suggest that TrpA1 is required cell-autonomously for light transduction in class IV dendritic arborization neurons.

Given the lack of AITC activation of class I or class III dendritic arborization neurons (Supplementary Fig. 17b), we expressed TrpA1 in class I dendritic arborization neurons and found that it conferred AITC sensitivity but not light response (Supplementary Fig. 17c, d), indicating that TrpA1 is not sufficient for light sensing. Because *trans*-heterozygotes carrying one mutant allele of *TrpA1* and one copy of the *MiET1* P-element insertion in the *Gr28b* gene showed reduced light response (Supplementary Fig. 19), it is likely that Gr28b and TrpA1 function in the same phototransduction pathway.

Class IV neurons mediate light avoidance behaviour

To test whether class IV dendritic arborization neurons are involved in light avoidance, we genetically ablated class IV dendritic arborization neurons of third instar larvae by expressing the pro-apoptotic genes *Hid* (ref. 15) and *reaper* (*rpr*) (ref. 37) via *ppk-GAL4* (*ppk-GAL4*; *UAS-Hid,rpr*) (Supplementary Fig. 1). We also constructed a line lacking Bolwig organs as well as class IV dendritic arborization neurons (*UAS-Hid,rpr*; *GMR-Hid*; *ppk-GAL4*). Notably, both lines showed markedly decreased white-light-avoidance behaviour compared to wild-type and *GMR-Hid* (Bolwig-organ-ablated) larvae (Fig. 5a, b). Class IV dendritic-arborization-neurons-ablated animals showed a significant decrease of avoidance versus wild type, for all white light intensities tested (Fig. 5c–g). Ablation of class IV dendritic arborization neurons in animals lacking Bolwig organs produced a further decrease in white light avoidance (Fig. 5d–g). Avoidance of high-intensity (>0.57 mW mm^{-2}) white light was normal when Bolwig organs were ablated in wild type (Fig. 5e–g), and in control strains with either *GAL4* or *UAS* (Fig. 5f). Taken together with similar
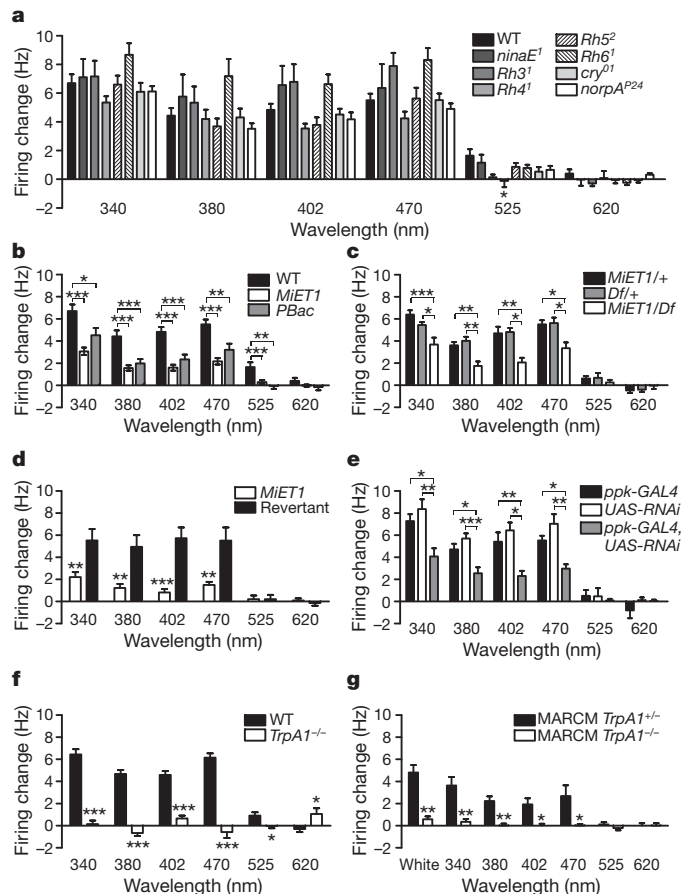
Figure 4 | Gr28b and TrpA1 are essential for class IV dendritic arborization neuron light responses. a, No significant defects were detected between wild-type and mutants of known phototransduction molecules with 340, 380, 402, 470, or 620 nm light. $n = 5$–10. **b,** Reduced light response of class IV dendritic arborization neurons in *MiET1* and *PBac* larvae. $n = 8$–29. **c,** Reduced light response of class IV dendritic arborization neurons in *MiET1*/deficiency larvae. $n = 5$–12. **d,** Precise excision of *MiET1* P-element insertion restores light response in class IV dendritic arborization neurons. $n = 6$–9. **e,** Reduced light responses of class IV dendritic arborization neurons with *Gr28b* RNAi knockdown. $n = 5$–8. **f,** Abolished light responses of class IV dendritic arborization neurons in *TrpA1*$^{-/-}$ mutants. $n = 8$–13. **g,** MARCM analysis of *TrpA1*$^{+/-}$ and *TrpA1*$^{-/-}$ class IV dendritic arborization neurons' response to light. $n = 5$–8. For **a–g**, Light intensities (mW mm^{-2}) are: 1.15 (340 nm), 5.79 (380 nm), 11.4 (402 nm), 52.8 (470 nm), 43.4 (525 nm), 29.6 (620 nm) and 94.7 (white). For **a, b, c, e**, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; one-way ANOVA followed by a Bonferroni post test; for **d, f, g**, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; two-tailed unpaired t-test. All error bars indicate s.e.m.

findings with ultraviolet, violet and blue light (Supplementary Figs 20–22), these results demonstrate that class IV dendritic arborization neurons are necessary to elicit photoavoidance at high intensities. It thus seems that the Bolwig organs and class IV dendritic arborization neurons operate in different light intensity regimes: Bolwig organs are tuned to low light, whereas class IV dendritic arborization neurons, required in low light, are the primary sensors at high intensities.

Careful examination of *ppk-GAL4* revealed additional expression in four mouth hook neurons, but not in the central nervous system (Supplementary Fig. 23a–d). Laser ablation of these four neurons in the *GMR-Hid* background had no effect on light avoidance behaviour (Supplementary Fig. 23e). Therefore, the class IV dendritic arborization neurons in the body wall are the ones important for the light avoidance behaviour.

Pickpocket, a Degenerin/Epithelial sodium Channel (DEG/ENaC) family member specifically expressed in class IV dendritic arborization neurons[24,38] (Fig. 3c), has been implicated in locomotion control[39–41].

Figure 5 | Class IV dendritic arborization neurons are the extra-ocular photoreceptors that contribute to light avoidance. a, b, Examples of larvae with either class IV dendritic arborization neurons ablated (**a**) or both Bolwig organs and class IV dendritic arborization neurons ablated (**b**) that failed to respond to white light (0.57 mW mm^{-2}) applied from 0 to 5 s (dotted circle). Arrow indicates locomotion direction. Scale bar, 1 mm (**a, b**), shown at −2 s. **c–g**, Percentage of animals avoiding white light of different intensities (in mW mm^{-2}: **c**, 0.088; **d**, 0.24; **e**, 0.57; **f**, 1.0; **g**, 1.67). Wild-type larvae, Bolwig-organ-ablated larvae (*GMR-Hid*), larvae with class IV dendritic arborization neurons ablated (*ppk-GAL4; UAS-Hid,rpr*) and larvae with both ablated (*UAS-Hid,rpr*; *GMR-Hid*; *ppk-GAL4*) were examined. **h**, Percentage of Bolwig-organ-ablated animals avoiding 0.25 mW mm^{-2} 525 nm green light when class IV dendritic arborization neurons express ChR2 with or without dietary retinal. **i**, Percentage of animals avoiding white light at 1 mW mm^{-2}. For **c–i**, controls are black bars. Twenty to forty animals were tested for each condition; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; two-tailed Fisher exact test. ChR2, channelrhodopsin-2; *rpr*, *reaper*; NS, not significant.
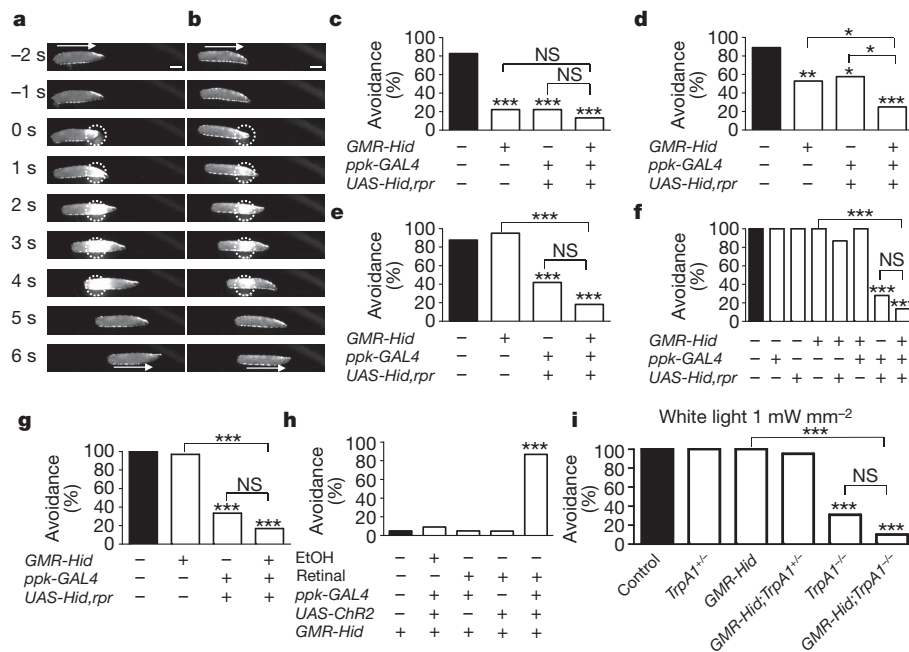
However, nose-touch experiments[42] revealed that larvae lacking class IV dendritic arborization neurons responded normally to gentle touch by retracting or turning away their heads (Supplementary Fig. 24). Moreover, direct recording of class IV dendritic arborization neurons in *ppk* mutant larvae revealed no defect in light response (Supplementary Fig. 15b). These results demonstrate that reduced light avoidance in class IV dendritic-arborization-ablated larvae is not due to non-specific effects.

To probe sufficiency, we expressed channelrhodopsin-2 (ChR2), a retinal-dependent cation channel gated by light from ultraviolet to green[43–45], specifically in class IV dendritic arborization neurons. ChR2 conferred green light sensitivity to dendritic arborization neurons from larvae fed with retinal (Supplementary Fig. 25), as well as robust avoidance of green light of retinal-fed larvae without Bolwig organs (Fig. 5h). Thus, activation of class IV dendritic arborization neurons is sufficient to induce avoidance.

With or without Bolwig organs, *TrpA1* mutant larvae showed deficient avoidance of 1 mW mm^{-2} white light (Fig. 5i). Moreover, reducing TrpA1 expression in class IV dendritic arborization neurons by RNAi was sufficient to abolish the light avoidance behaviour in animals without Bolwig organs (Supplementary Fig. 26). Together, our physiological and behavioural studies indicate that a light transduction pathway involving TrpA1 and Gr28b in class IV dendritic arborization neurons is necessary for light avoidance.

Discussion

Extra-ocular photoreceptors, previously found in reptiles, birds, amphibians and fish, provide a good measure of ambient light luminance and serve mainly non-image-forming functions such as phototaxis, circadian photo-entrainment, pupal reflex, shadow reaction and magnetic orientation[1–7]. Usually, these extra-ocular photoreceptors have much lower light sensitivity and slower kinetics than ocular photoreceptors[3].

Drosophila larvae have primitive eye structures, the Bolwig organs, which control avoidance of dim light[12]. Here we report that the class IV dendritic arborization neurons, previously implicated in mechanosensory response and motion control[38–41,46], are surprisingly also photoreceptors. Our behavioural analysis suggests that Bolwig organs and class IV dendritic arborization neurons have different regimes of light sensing in acute photoavoidance. Bolwig organs, packed with photopigments[47], are preferentially required for avoidance of low light. Class IV dendritic arborization neurons, which also contribute to low light avoidance, are the primary sensors at sunlight-level intensities.

This organization ensures that larvae can detect the full range of ambient light intensities, from dim to strong.

Class IV dendritic arborization neurons have the intrinsic ability to sense light, even after isolation in culture, and their dendrites are capable of sensing light (Fig. 3 and Supplementary Fig. 10). Importantly, the dendrites of class IV dendritic arborization neurons have complete and non-redundant coverage of the body wall (Fig. 3c), allowing animals to perceive illumination of any body part, and initiate an appropriate behavioural response. Larvae spend much of the time with their heads digging into food, making their Bolwig organs on the head less likely to be exposed to light. Thus, the ability to sense light with sensory neurons tiling the body wall is critical for detection of exposure.

Class IV dendritic arborization neurons use a novel light transduction pathway. Like in *C. elegans*, a putative chemosensory G-protein-coupled receptor, Gr28b, is involved for phototransduction in class IV dendritic arborization neurons (Fig. 4b–e). TrpA1 also is essential (Fig. 4f, g and Supplementary Fig. 18). *Drosophila* larval class IV dendritic arborization neurons may function as nociceptors[46,48,49]. They are required for thermal and mechanical nociception, and activation of class IV dendritic arborization neurons is sufficient to induce a behaviour pattern similar to nocifension[46,48,49]. Given that class IV dendritic arborization neurons are required for larvae to avoid harmful light stimuli, these neurons seem to be poised to alert the animal to a variety of adversities.

Our study has uncovered unexpected light-sensing machinery, which could be critical for foraging larvae to avoid harmful sunlight, desiccation and predation. By providing precedence for photoreceptors strategically placed away from the eyes, our finding of an array of class IV dendritic arborization neurons with elaborate dendrites tiling the entire body wall, and acting as light-sensing antennae, raises the question of whether other animals with eyes might also possess extra-ocular photoreceptors for more thorough light detection and behavioural response.

METHODS SUMMARY

Light avoidance assay. Light avoidance was scored if the third instar larva reversed in direction or turned its head completely away from the 1.7-mm light spot on its head during the 5-s illumination. Two-tailed Fisher exact test (20−40 larvae per condition), *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Electrophysiology. Action potentials were monitored via extracellular recordings from a third instar larval fillet with muscles removed, using an Axon 700B amplifier and pCLAMP 10 software.

GCaMP3 imaging. Third instar larval fillets were imaged on a Zeiss LS510 META confocal microscope with an Olympus ×40/0.8 NA water immersion objective, with a 488-nm laser. GCaMP3 cDNA is available from AddGene.

Cell culture. Embryos homozygous for Canton S (*Cs*); *UAS-GCaMP3; ppk-GAL4, UAS-RFP* (for class IV dendritic arborization neuron culture) or *Cs*; *UAS-GCaMP3; 19-12-GAL4, UAS-RFP* (for class III dendritic arborization neuron culture) were used for culture[22,23].

MARCM analysis. We recorded from class IV dendritic arborization neuron clones marked with GFP for lacking *TrpA1* (ref. 36).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Fu, Y., Liao, H. W., Do, M. T. & Yau, K. W. Non-image-forming ocular photoreception in vertebrates. *Curr. Opin. Neurobiol.* **15,** 415–422 (2005).
2. Yau, K. W. & Hardie, R. C. Phototransduction motifs and variations. *Cell* **139,** 246–264 (2009).
3. Steven, D. M. The dermal light sense. *Biol. Rev. Camb. Philos. Soc.* **38,** 204–240 (1963).
4. Millott, N. The dermal light sense. *Symp. Zool. Soc. Lond.* **23,** 1–36 (1968).
5. Yoshida, M. Extraocular photoreception. In *Handbook of Sensory Physiology* Vol. 7/6A, 581–640 (Springer, 1979).
6. Halford, S. *et al.* VA opsin-based photoreceptors in the hypothalamus of birds. *Curr. Biol.* **19,** 1396–1402 (2009).
7. Phillips, J. B., Deutschlander, M. E., Freake, M. J. & Borland, S. C. The role of extraocular photoreceptors in newt magnetic compass orientation: parallels between light-dependent magnetoreception and polarized light detection in vertebrates. *J. Exp. Biol.* **204,** 2543–2552 (2001).
8. Ward, A., Liu, J., Feng, Z. & Xu, X. Z. Light-sensitive neurons and channels mediate phototaxis in *C. elegans. Nature Neurosci.* **11,** 916–922 (2008).
9. Edwards, S. L. *et al.* A novel molecular solution for ultraviolet light detection in *Caenorhabditis elegans. PLoS Biol.* **6,** e198 (2008).
10. Liu, J. *et al. elegans* phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog. *Nature Neurosci.* **13,** 715–722 (2010).
11. Sawin-McCormack, E. P., Sokolowski, M. B. & Campos, A. R. Characterization and genetic analysis of *Drosophila melanogaster* photobehavior during larval development. *J. Neurogenet.* **10,** 119–135 (1995).
12. Mazzoni, E. O., Desplan, C. & Blau, J. Circadian pacemaker neurons transmit and modulate visual information to control a rapid behavioral response. *Neuron* **45,** 293–300 (2005).
13. Sprecher, S. G. & Desplan, C. Switch of rhodopsin expression in terminally differentiated *Drosophila* sensory neurons. *Nature* **454,** 533–537 (2008).
14. Willson, R. C., Gulkis, S., Janssen, M., Hudson, H. S. & Chapman, G. A. Observations of solar irradiance variability. *Science* **211,** 700–702 (1981).
15. Grether, M. E., Abrams, J. M., Agapite, J., White, K. & Steller, H. The head involution defective gene of *Drosophila melanogaster* functions in programmed cell death. *Genes Dev.* **9,** 1694–1708 (1995).
16. Hay, B. A., Maile, R. & Rubin, G. M. P element insertion-dependent gene activation in the *Drosophila* eye. *Proc. Natl Acad. Sci. USA* **94,** 5195–5200 (1997).
17. Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19,** 137–141 (2001).
18. Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112,** 271–282 (2003).
19. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6,** 875–881 (2009).
20. Ueda, A. & Wu, C. F. Effects of *hyperkinetic*, a β subunit of *Shaker* voltage-dependent K^+ channels, on the oxidation state of presynaptic nerve terminals. *J. Neurogenet.* **22,** 103–115 (2008).
21. Auld, V. J., Fetter, R. D., Broadie, K. & Goodman, C. S. Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in *Drosophila. Cell* **81,** 757–767 (1995).
22. Saito, M. & Wu, C. F. Expression of ion channels and mutational effects in giant *Drosophila* neurons differentiated from cell division-arrested embryonic neuroblasts. *J. Neurosci.* **11,** 2135–2150 (1991).
23. Bai, J., Sepp, K. J. & Perrimon, N. Culture of *Drosophila* primary cells dissociated from gastrula embryos and their use in RNAi screening. *Nature Protocols* **4,** 1502–1512 (2009).
24. Grueber, W. B. *et al.* Projections of *Drosophila* multidendritic neurons in the central nervous system: links with peripheral dendrite morphology. *Development* **134,** 55–64 (2007).
25. Grueber, W. B., Jan, L. Y. & Jan, Y. N. Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. *Development* **129,** 2867–2878 (2002).
26. Grueber, W. B., Ye, B., Moore, A. W., Jan, L. Y. & Jan, Y. N. Dendrites of distinct classes of *Drosophila* sensory neurons show different capacities for homotypic repulsion. *Curr. Biol.* **13,** 618–626 (2003).
27. O'Tousa, J. E. *et al.* The *Drosophila ninaE* gene encodes an opsin. *Cell* **40,** 839–850 (1985).
28. Dolezelova, E., Dolezel, D. & Hall, J. C. Rhythm defects caused by newly engineered null mutations in *Drosophila*'s *cryptochrome* gene. *Genetics* **177,** 329–345 (2007).
29. Montell, C. Visual transduction in *Drosophila. Annu. Rev. Cell Dev. Biol.* **15,** 231–268 (1999).
30. Thorne, N. & Amrein, H. Atypical expression of *Drosophila* gustatory receptor genes in sensory and central neurons. *J. Comp. Neurol.* **506,** 548–568 (2008).
31. Montell, C., Jones, K., Hafen, E. & Rubin, G. Rescue of the *Drosophila* phototransduction mutation trp by germline transformation. *Science* **230,** 1040–1043 (1985).
32. Rosenzweig, M. *et al.* The *Drosophila* ortholog of vertebrate TRPA1 regulates thermotaxis. *Genes Dev.* **19,** 419–424 (2005).
33. Hamada, F. N. *et al.* An internal thermal sensor controlling temperature preference in *Drosophila. Nature* **454,** 217–220 (2008).
34. Kwon, Y., Shim, H. S., Wang, X. & Montell, C. Control of thermotactic behavior via coupling of a TRP channel to a phospholipase C signaling cascade. *Nature Neurosci.* **11,** 871–873 (2008).
35. Kang, K. *et al.* Analysis of *Drosophila* TRPA1 reveals an ancient origin for human chemical nociception. *Nature* **464,** 597–600 (2010).
36. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).
37. White, K. *et al.* Genetic control of programmed cell death in *Drosophila. Science* **264,** 677–683 (1994).
38. Adams, C. M. *et al.* Ripped pocket and pickpocket, novel *Drosophila* DEG/ENaC subunits expressed in early development and in mechanosensory neurons. *J. Cell Biol.* **140,** 143–152 (1998).
39. Ainsley, J. A., Kim, M. J., Wegman, L. J., Pettus, J. M. & Johnson, W. A. Sensory mechanisms controlling the timing of larval developmental and behavioral transitions require the *Drosophila* DEG/ENaC subunit, Pickpocket1. *Dev. Biol.* **322,** 46–55 (2008).
40. Ainsley, J. A. *et al.* Enhanced locomotion caused by loss of the *Drosophila* DEG/ENaC protein Pickpocket1. *Curr. Biol.* **13,** 1557–1563 (2003).
41. Xu, K. *et al.* The *fragile X-related* gene affects the crawling behavior of *Drosophila* larvae by regulating the mRNA level of the DEG/ENaC protein Pickpocket1. *Curr. Biol.* **14,** 1025–1034 (2004).
42. Kernan, M., Cowan, D. & Zuker, C. Genetic dissection of mechanosensory transduction: mechanoreception-defective mutations of *Drosophila. Neuron* **12,** 1195–1206 (1994).
43. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100,** 13940–13945 (2003).
44. Schroll, C. *et al.* Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in *Drosophila* larvae. *Curr. Biol.* **16,** 1741–1747 (2006).
45. Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. & Deisseroth, K. Circuit-breakers: optical technologies for probing neural signals and systems. *Nature Rev. Neurosci.* **8,** 577–581 (2007).
46. Zhong, L., Hwang, R. Y. & Tracey, W. D. Pickpocket is a DEG/ENaC protein required for mechanical nociception in *Drosophila* larvae. *Curr. Biol.* **20,** 429–434 (2010).
47. Pollock, J. A. & Benzer, S. Transcript localization of four opsin genes in the three visual organs of *Drosophila*; RH2 is ocellus specific. *Nature* **333,** 779–782 (1988).
48. Tracey, W. D. Jr, Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113,** 261–273 (2003).
49. Hwang, R. Y. *et al.* Nociceptive neurons protect *Drosophila* larvae from parasitoid wasps. *Curr. Biol.* **17,** 2105–2116 (2007).

Author Contributions Y.X. designed and carried out the experiments and analysed the data; Q.Y. characterized molecular information of Gr28b, rhodopsin and cryptochrome. L.L.L. created GCaMP3 and did the bioinformatic analyses of Gr28b; N.V. cleaned up the *Rh3^1* and *Rh4^1* mutants; Y.N.J. helped to design the experiments and supervised the work; Y.X., L.L.L., L.Y.J. and Y.N.J. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.N.J. (yuhnung.jan@ucsf.edu).

METHODS

Fly stocks. The following fly strains were used: (1) *Cs*; (2) *Cs; GMR-Hid*; (3) *Cs; ppk-GAL4, UAS-Tomato*; (4) *Cs;; ppk-GAL4, UAS-mCD8::GFP*; (5) *Cs;; ppk-GAL4, UAS-mCD8::RFP*; (6) *Cs;; TrpA1*; (7) *w; UAS-TrpA1 RNAi*; (8) *Cs; UAS-Dicer; ppk-GAL4, UAS-mCD8::GFP*; (9) *Cs; 21-7-GAL4, UAS-GCaMP3*; (10) *Cs; UAS-GCaMP3; ppk-GAL4, UAS-mCD8::RFP*; (11) *Cs; UAS-GCaMP3; 19-12-GAL4, UAS-mCD8::RFP*; (12) *elav-GAL4, hsFLP, UAS-mCD8::GFP;;tub-GAL80, FRT^{2A}*; (13) *w;; TrpA1, FRT^{2A}/TM6B, Tb*; (14) *Cs; MiET103888* (Bloomington stock centre no. 24190); (15) *Cs; PBac01884* (Bloomington stock centre no. 10743); (16) *w; UAS-Gr28b RNAi* (VDRC stock centre no. kk101727); (17) *Cs;; cry^{01}* (ref. 28); (18) *Cs;; Rh3^1*; (19) *Cs;; Rh4^1*; (20) *Cs; Rh5^2*; (21) *Cs;; Rh6^1*; (22) *ninaE^{17}*; (23) *norpA^{P24}*; (24) *Df(2L)Exel7031*; (25) *yw, UAS-Hid,rpr*; (26) *Cs; Channelrhodopsin-2*; (27) *UAS-TrpA1*; (28) *Cs; trpl^{302}*; (29) *Cs; painless^1*.

Light avoidance assay. Animals were raised at 25 °C in an incubator with 12 h light/dark cycles and humidity control (Darwin Chamber Company). Ninety-six hours after egg laying (AEL), third instar larvae were gently picked up from the vial, washed twice with PBS and transferred to a 100-mm Petri dish with fresh 2% agarose. Excessive water was removed from the animals. Animals were allowed to rest on the plate for at least 3 min before testing. Only animals making straight forward movement were selected for the assay. Each animal was tested once. The assay was carried out with a Stereo Microscope system (Leica M205FA). Unless otherwise specified, light was delivered from a 300 W xenon lamp (Sutter LB-LS/30) through a PLANAPO ×1 objective (Leica) at ×160 magnification, yielding a light spot of 1.7 mm in diameter. To direct the light to the animal's head, the plate was manually moved so that only the head appeared in the field of view. An avoidance response was scored when animals stopped forward movement during the 5-s light illumination by either initiating backward movement or turning their heads completely away from the light spot. The 5-s light illumination was controlled by a shutter (Sutter Instruments) in the xenon lamp house triggered by an external stimulator (Grass s88). The xenon lamp has a light intensity spectrum similar to sunlight. The background light for visualizing animals was filtered to red, to which they are insensitive (Lee filter no. 027, medium red), and the entire event was recorded through a lens (Fujinon, 25 mm 1:1.4) to a CCD camera (Qimaging Rolera XR) at 6 frames per second. The camera was mounted on a tripod and placed on the side of the Petri dish, with the front of the lens covered with red filters (Lee filter no. 027, medium red) to avoid overexposure. For violet, blue, green or red light, the band-pass excitation filter (in nm: 402 ± 7.5, 470 ± 20, 525 ± 25, 620 ± 30) was placed in the xenon lamp house, and a filter set with empty excitation filter was placed into the Leica scope. White light illumination was achieved the same way except that no excitation filter was placed in the xenon lamp house. For 360 nm illumination, a HXP-120 light source (Visitron Systems) was used, and a 360 ± 20 nm excitation filter set was placed into the Leica scope. The light intensity at ×160 magnification was measured by a radiometric sensor head (Newport 818P-001-12) coupled with a power meter (Newport 1918-C). Because liquid light guides connecting the light source and the microscope ensured the uniformity of light, the light intensity was calculated by dividing the measured light intensity over area (2.27 mm^2). Temperature changes associated with light illumination were measured with an IT-24P thermocouple probe coupled with a BAT-10 thermometer (Physitemp). A red filter (Lee filter no. 027, medium red) was used to cover the eyepieces of the microscope to protect experimenters' eyes from strong light. Light avoidance of animals expressing channelrhodopsin-2 (ChR2) to 0.25 mW mm^{-2} green light was done the same way with the exception that eggs were laid and allowed to develop to third instar in food medium supplemented with 0.2 mM retinal. Twenty to forty animals were tested in each condition and the percentage of positive responses was calculated. A two-tailed Fisher exact test was performed and statistical significance was assigned, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Cell culture. Dissociated cell cultures were prepared from early gastrulas of *Drosophila melanogaster*, as described previously[22,23]. Briefly, embryos were collected on grape agar plates with yeast paste and incubated for another 3.5 h at 25 °C. After removing yeast paste carefully, embryos were washed extensively with 500 ml sterile H$_2$O. To remove the chorionic membrane and sterilize the embryos, embryos were treated with bleach/90% EtOH (1:1 by volume, final concentration of sodium hypochlorite is 3%) solution for 1 min and then washed with 500 ml sterile H$_2$O to remove residual bleach and EtOH. After wash, embryos were homogenized in Schneider's medium supplemented with 5% FBS, 0.2 µg ml^{-1} insulin, penicillin (50 units ml^{-1}) and streptomycin (50 µg ml^{-1}), with a 15-ml Dounce homogenizer containing 6 ml medium. Three to five rotary up and down strokes were used to dissociate the cells. To remove undissociated clumps and large debris, dissociated cells were filtered through a 40-µm cell strainer. Filtrate was collected in a 15-ml centrifuge tube, and cells were centrifuged at 2,000 r.p.m. for 5 min. Cells were washed in the medium and centrifuged again as described, followed by a final suspension in 10 ml medium.

Cells were plated on poly-L-lysine-treated coverglass, and allowed to develop at room temperature for 2–4 days before GCaMP3 imaging was performed. Embryos carrying *Cs; UAS-GCaMP3; ppk-GAL4, UAS-mCD8::RFP* were used for culturing class IV dendritic arborization neurons, and embryos carrying *Cs; UAS-GCaMP3; 19-12-GAL4, UAS-mCD8::RFP* were used for culturing class III dendritic arborization neurons.

Electrophysiology. Fillets were made from 96-h-AEL third instar larvae with cuticle facing down in the external saline solution composed of (in mM): NaCl 120, KCl 3, MgCl$_2$ 4, CaCl$_2$ 1.5, NaHCO$_3$ 10, trehalose 10, glucose 10, TES 5, sucrose 10, HEPES 10. Osmolality was 305 mOsm kg^{-1} and final pH = 7.25. Muscles covering the neurons of interest were gently digested by proteinase (Sigma). During the enzymatic treatment, a laminar flow of external saline solution was turned on to remove excess enzyme. No detectable difference in dendrite morphology of class IV dendritic arborization neurons was observed before and after muscle digestion, indicating that neurons were intact. An Olympus BX51WI microscope with a ×40/0.8 NA water immersion objective was used to obtain recordings with the help of IR-DIC optics and a CoolSNAP CCD (Photometrics). Recording pipettes were pulled with P-97 puller (Sutter instruments) from thin wall borosilicate glass (World Precision Instruments), filled with external saline solution, with a tip opening of 5 µm. Gentle negative pressure was delivered to suck the soma to get good signal-to-noise ratio of recording traces. Recordings were performed with a Multiclamp 700B amplifier (Molecular Devices), and data were acquired with Digidata 1440A (Molecular Devices) and Clampex 10.0 software (Molecular Devices). Extracellular recordings of action potentials were obtained in voltage clamp mode with a holding potential of 0 mV, with a 2 kHz low-pass filter and sampled at 20 kHz. During recording, no background light illumination was applied. A 300-W xenon light source was connected to the microscope with a liquid light guide to provide light stimulation through a ×40/0.8 NA water-immersion lens, yielding an evenly illuminated light spot with 600 µm diameter, which covered the entire class IV dendritic arborization neuron. Dendritic illumination was achieved by decreasing the field diaphragm to cover about 50% of dendrites (Fig. 3d), without illuminating the soma. Light intensity was measured the same way as in the behaviour assay and intensity density was calculated. Neutral density filters (Chroma) were used to reduce light intensity to generate dose-response curves. The duration (5 s) of light illumination was controlled by a shutter in the xenon lamp house triggered by Digidata 1440A (Molecular Devices). Band-pass excitation filters (Chroma) were used to select light wavelength: they were (in nm) 340 ± 10, 380 ± 10, 402 ± 7.5, 470 ± 20, 525 ± 25, 620 ± 30. White light illumination was achieved the same way with the exception that no excitation filter was placed in the xenon lamp house. For each recording trace, average frequency during the 5 s immediately before light exposure was used as control. Five-second light stimulation was controlled by a TTL-triggered shutter (Sutter Instruments) in the xenon lamp house. For latency analysis, only neurons with low spontaneous firing were included for recording, and latency was defined as the time between onset of light and onset of burst firing. To record temperature-induced firing change, preheated solution was perfused into the recording chamber. Temperature was monitored by the thermal probe connected with the thermometer (Warner Tc-324B). mSIRK (EMD bioscience) or L-*cis*-diltiazem (Sigma) was incubated in the recording chamber for 30 min before recording. A two-tailed paired or unpaired *t*-test, or one-way ANOVA, followed by the Bonferroni multiple comparison test, was performed and statistical significance was assigned, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

GCaMP3 imaging experiments. Homozygous *Cs; 21-7-GAL4, UAS-GCaMP3* animals were used for imaging. The *21-7* promoter drives Gal4 expression in all peripheral nervous system sensory neurons in the larval body wall except the chordotonal organ (H. H. Lee, Y.X., L.Y.J. and Y.N.J., unpublished data). Fillet preparation was the same as the one used in recording with the exception that the cuticle was facing up to mimic the orientation of larvae in receiving natural light. No enzymatic digestion was performed. Data were collected on a Zeiss LS510 META confocal microscope with an Olympus ×40/0.8 NA water immersion objective. GCaMP3 fluorescence was excited with a 488-nm laser[18,19]. Laser scanning by itself didn't activate class IV dendritic arborization neurons, as evidenced by the flat baseline of GCaMP3 signals during the 0–60 s control period (Fig. 2e). The images were acquired at 512 × 512 pixels at 12-bit dynamic range. The duration of 5 s of light stimulation (91, 68 and 96 mW mm^{-2} for 365 nm ultraviolet, 470 nm blue and 546 nm green light, respectively) was controlled by manually switching the filter cube from laser-scanning position to epifluorescence position. Two seconds of light stimulation elicited similar results as five seconds (data not shown). Average GCaMP3 signals from 60 s before light stimulation was taken as F_0, and $\Delta F/F_0$ was calculated for each data point. GCaMP3 signals from the soma were analysed, although axons and dendrites also showed responses. As a control, *Cs; UAS-GCaMP3; ppk-GAL4, UAS-mCD8::RFP* animals

were raised to look for nonspecific effects of excitation. In these animals, we did not detect any change in RFP signals in response to light (data not shown).

For GCaMP3 imaging of cultured neurons, class IV or class III dendritic arborization neurons were identified by the co-expressed RFP signals. Light stimulation was carried out as described above. For AITC stimulation, an equal volume of 200 μM AITC was manually applied to the chamber, making the final concentration 100 μM.

MARCM analysis. MARCM analysis was carried out as described previously[36]. *TrpA1* mutant class IV dendritic arborization neurons with GFP signals were selected for recording. For $TrpA1^{+/-}$ heterozygous neurons without GFP expression, class IV dendritic arborization neurons were identified by location.

Nose-touch assay. Assay was performed as described previously[42]. Briefly, the larvae were touched with an eyebrow hair affixed to the tip of a dissecting needle. The scoring system is as follows: 0 = no response to touch; 1 = a response of pausing mouth hook movement; 2 = responding by withdrawing the anterior or turning away from the touch; 3 = a single reverse peristaltic wave away from the touch; and 4 = multiple peristaltic waves away from the touch. A two-tailed Fisher exact test was performed.

LETTER

Growth of graphene from solid carbon sources

Zhengzong Sun[1], Zheng Yan[1], Jun Yao[2], Elvira Beitler[1], Yu Zhu[1] & James M. Tour[1,3]

Monolayer graphene was first obtained[1] as a transferable material in 2004 and has stimulated intense activity among physicists, chemists and material scientists[1–4]. Much research has been focused on developing routes for obtaining large sheets of monolayer or bilayer graphene. This has been recently achieved by chemical vapour deposition (CVD) of CH_4 or C_2H_2 gases on copper or nickel substrates[5–7]. But CVD is limited to the use of gaseous raw materials, making it difficult to apply the technology to a wider variety of potential feedstocks. Here we demonstrate that large area, high-quality graphene with controllable thickness can be grown from different solid carbon sources—such as polymer films or small molecules—deposited on a metal catalyst substrate at temperatures as low as 800 °C. Both pristine graphene and doped graphene were grown with this one-step process using the same experimental set-up.

With its extraordinary electronic and mechanical properties, graphene is showing promise in a plethora of applications[7–13]. Graphene can now be obtained by several different approaches. The original mechanical peeling method from highly oriented pyrolytic graphite yields small amounts of high quality graphene[1]. Liquid exfoliation and reduction of graphene oxide have been used to produce chemically converted graphene in large quantities[14,15]. Annealing SiC, growth from amorphous carbon and CVD methods have been used to synthesize large-size graphene on wafers[5–7,16,17]. By introducing Ni and Cu as the substrates for CVD growth, the size, thickness and quality of the produced graphene is approaching industrially useful specifications[5–7]. However, intrinsic graphene is a zero bandgap material that shows a weak ambipolar behaviour; transistors based on such graphene show small 'on'/'off' current ratios, so they are too metallic for many designed electronics applications[18]. In order to modify the Fermi level of graphene and manipulate its electronic and optical properties, doping the graphene matrix with heteroatoms is a straightforward way to make an n-type, p-type or hybrid doped graphene[19–23].

In the present work, the growth of monolayer pristine graphene from solid carbon sources atop metal catalysts is demonstrated (Fig. 1a). The first solid carbon source used was a spin-coated poly(methyl methacrylate) (PMMA) thin film (~100 nm) and the metal catalyst substrate was a Cu film. At a temperature as low as 800 °C or as high as 1,000 °C (tested limit) for 10 min, with a reductive gas flow (H_2/Ar) and under low pressure conditions, a single uniform layer of graphene was formed on the substrate. The graphene material thus produced was successfully transferred to different substrates for further characterization (see Supplementary Materials and Supplementary Methods).

The Raman spectrum of this monolayer PMMA-derived graphene is shown in Fig. 1b and the spectrum is characteristic of >10 locations recorded over 1 cm^2 of the sample. The two most pronounced peaks in this spectrum are the G peak at 1,580 cm^{-1} and the 2D peak at 2,690 cm^{-1}. The I_{2D}/I_G intensity ratio is about 4 and the full-width at half-maximum of the 2D peak is about 30 cm^{-1}, indicating that the graphene is a monolayer. The D peak (~1,350 cm^{-1}) is in the noise level, indicating the presence of few sp^3 carbon atoms or defects[24].
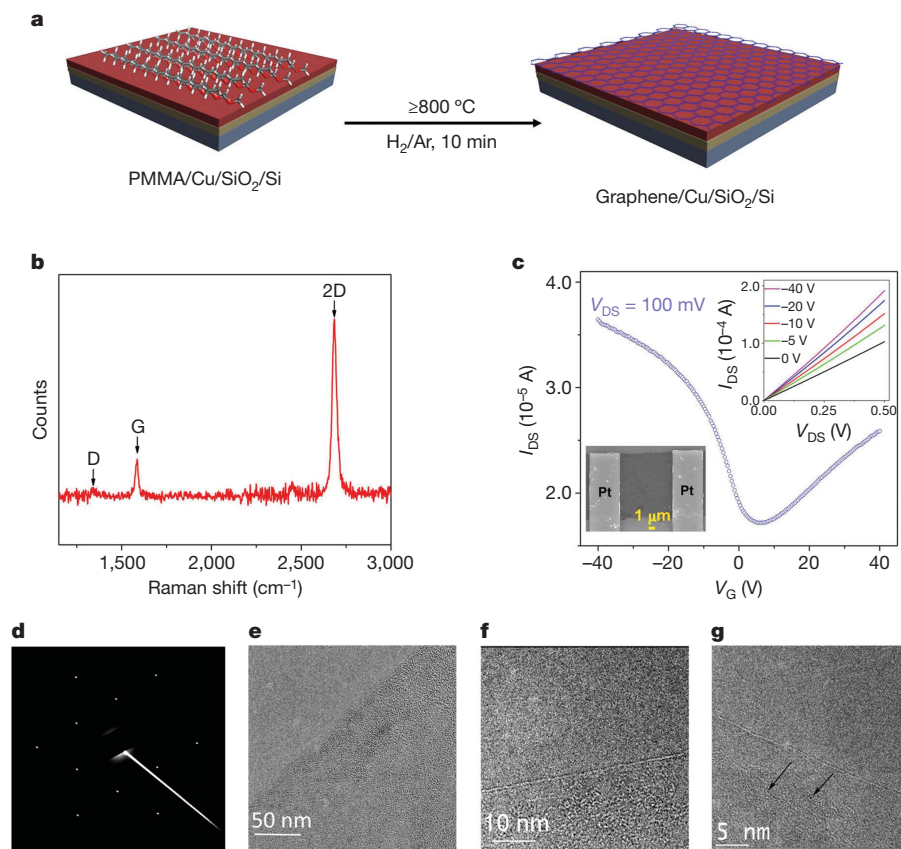
The electrical properties of the PMMA-derived graphene were evaluated with back-gated graphene-based field-effect transistor (FET) devices atop a 200-nm-thick SiO_2 dielectric. Typical data for the FET devices are shown in Fig. 1c. For this particular device, the estimated carrier (hole) mobility is ~410 $cm^2 V^{-1} s^{-1}$ at room temperature and the 'on'/'off' current ratio is ~2, which is expected in graphene-based FET devices of this size[21]. Although the graphene was pristine without any doping atoms, it still shows weak p-type behaviour, with the neutrality point moved to positive gate voltage; this probably arises from the physisorption of small molecules, such as H_2O (ref. 10). Placing graphene FETs under high vacuum (10^{-5}–10^{-6} torr) for several days moves the neutrality point to zero (Supplementary Fig. 1), confirming that the weak p-type behaviour was due to physisorption of volatile molecules[10].

Transmission electron microscopy (TEM) images of the pristine PMMA-derived graphene and its diffraction pattern are shown in Fig. 1d–g. The selected area electron diffraction (SAED) pattern in Fig. 1d displays the typical hexagonal crystalline structure of graphene. The layer count on the edges of the images indicates the thickness of this PMMA-derived graphene. The edges in Fig. 1e–g were randomly imaged under TEM and most were monolayer or bilayer graphene, which corroborates the Raman data. Although most of the graphene surface was continuous and crystalline according to its diffraction pattern, there is adsorbed PMMA resulting from the transfer step. Metal atoms or ions were also found to be trapped on the graphene surface (black arrows in Fig. 1g) and became charge impurities, which should increase the charge density but decrease the mobility of the PMMA-derived graphene[25]. Similar phenomena have been observed with CVD-generated graphene[5–7]. Atomic force microscopy was used to characterize the surface profile of PMMA-derived graphene on a SiO_2/Si substrate. In Supplementary Fig. 2, the thickness of this graphene is about 0.7 nm, which confirms the monolayer nature of this material. However, limited by the wet-transfer technique, graphene's intrinsic corrugation is still obvious in the AFM image.

Graphene's electronic properties are strongly linked to its thickness[26]. Therefore, it would be useful to be able to control the thickness when producing the graphene by tuning the growth parameters. We have found that the thickness of PMMA-derived graphene can be controlled—to give a monolayer, a bilayer, or a few layers—by changing the Ar and H_2 gas flow rate. Typical thicknesses were evaluated by Raman spectroscopy (Fig. 2a) and ultraviolet transmittance (Fig. 2b) of the graphene. At 1,000 °C, bilayer PMMA-derived graphene was obtained when the Ar flow rate was 500 cm^3 STP min^{-1} and the H_2 flow rate was 10 cm^3 STP min^{-1}. When the Ar flow rate was 500 cm^3 STP min^{-1} and the H_2 flow rate was 3–5 cm^3 STP min^{-1}, few-layer PMMA-derived graphene formed. When the H_2 flow rate was increased to 50 cm^3 STP min^{-1} or higher, only monolayer graphene was formed on the Cu substrate. Monolayer graphene showed a transmittance of 97.1% at 550 nm wavelength (Fig. 2b). It had a sheet resistance (R_s) of 1,200 Ω per square by the four-probe method, which makes it a transparent electrode material of interest. The bilayer graphene's transmittance at 550 nm wavelength is 94.3%, which shows linear enhancement in the ultraviolet absorption. The few-layer PMMA-derived graphene sheet in Fig. 2a has a transmittance of 83% at 550 nm, leading to an estimated six-layer thickness. Both the

[1]Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas 77005, USA. [2]Applied Physics Program, Department of Bioengineering, Rice University, 6100 Main Street, Houston, Texas 77005, USA. [3]Richard E. Smalley Institute for Nanoscale Science and Technology, Department of Mechanical Engineering and Materials Science, Rice University, 6100 Main Street, Houston, Texas 77005, USA.

Figure 1 | Synthetic protocol, spectroscopic analysis and electrical properties of PMMA-derived graphene. a, Monolayer graphene is derived from solid PMMA films on Cu substrates by heating in an H_2/Ar atmosphere at 800 °C or higher (up to 1,000 °C). **b**, Raman spectrum (514 nm excitation) of monolayer PMMA-derived graphene obtained at 1,000 °C. See text for details. **c**, Room temperature I_{DS}–V_G curve from a PMMA-derived graphene-based back-gated FET device. Top inset, I_{DS}–V_{DS} characteristics as a function of V_G; V_G changes from 0 V (bottom) to −40 V (top). Bottom inset, SEM (JEOL-6500 microscope) image of this device where the PMMA-derived graphene is perpendicular to the Pt leads. I_{DS}, drain–source current; V_G, gate voltage; V_{DS}, drain–source voltage. **d**, SAED pattern of PMMA-derived graphene. **e–g**, HRTEM images of PMMA-derived graphene films at increasing magnification. In **g**, black arrows indicate Cu atoms.

shape and the positions of the 2D peak are significantly different from monolayer graphene to bilayer graphene and few-layer graphene (Supplementary Fig. 3). For monolayer graphene, the 2D peak can be fitted with a single sharp Lorentz peak. The observed 2D splitting in bilayer and few-layer PMMA-derived graphene can be assigned to the electronic band splitting caused by the interaction of the graphene planes[24]. The Raman mapping of the G to 2D peak ratio illustrates the uniformity of the graphene films over the 70–75 μm^2 areas investigated (Supplementary Fig. 4). For monolayer graphene, more than 95% of the film has this signature, with $I_G/I_{2D} < 0.4$. The bilayer graphene has more than 85% coverage, with an $I_G/I_{2D} \approx 0.8$.

We interpret the effect of hydrogen as follows: hydrogen acts as both the reducing reagent and a carrier gas to remove C atoms that are extruded from the decomposing PMMA during growth. A slower H_2 flow leaves more C sources for the growth of multilayer graphene. Owing to the low concentration and solubility of the carbon source in Cu, traditional CVD-grown graphene on Cu usually terminates as a monolayer. In this experiment, highly concentrated and uniformly dispersed carbon sources favour multilayer graphene when the H_2 flow is low. The higher-order layers might form through graphitization directly atop the first layer, which blocks the contact of the carbon sources with the metal catalyst. Some metal catalysts, such as Ni, are known to reverse graphene growth by converting graphene to hydrocarbon products, therefore cutting graphene along specific directions[27]. This reverse reaction does not appear to occur on the PMMA-derived graphene which is atop the Cu.

High quality monolayer PMMA-derived graphene was obtained at 800 °C by this method; this is lower than the CVD growth temperature on Cu used in the original report[6] (see Supplementary Fig. 5). For the semiconductor industry, the lower processing temperature is favourable because temperatures as high as 1,000 °C would be problematic in the fabrication of the multi-layer stacks of heterogeneous materials. Therefore, in addition to changing the Ar/H_2 flow rate, the graphene growth process was conducted at different temperatures. The quality of the graphene films was monitored by the D/G peak ratio from Raman spectroscopic analysis. The peak ratio for graphene sheets obtained at 800 °C was less than 0.1. At 750 °C, the peak ratio was ~0.35; hence 800 °C is the lower limit for obtaining high quality graphene from PMMA (Supplementary Fig. 5). We used other solid carbon sources—including fluorene ($C_{13}H_{10}$) and sucrose (table sugar, $C_{12}H_{22}O_{11}$)—to grow monolayer graphene on Cu catalyst under the same growth conditions as was used for the PMMA-derived graphene. Because these precursors are powders not films, 10 mg of each as a finely ground powder was placed directly on a 1 cm^2 Cu foil. After subjecting the powder-coated Cu films to the same reaction conditions as used for PMMA-derived graphene, Raman spectra indicated that all of the solid carbon sources were transformed into monolayer graphene with no D peak observed (Fig. 2c). Although these solid carbon precursors contain potential topological defect generators (the five-membered ring in fluorene) or high concentrations of heteroatoms (oxygen in sucrose), they produce high quality pristine graphene. It is possible that at elevated temperatures under vacuum, C has a higher affinity for the metal catalyst surface than the heteroatoms; atom rearrangement occurs and most of the topological defects are self-healed as the graphene is formed.

Other substrates—such as Ni, Si<100> with native oxide, and 200-nm-thick SiO$_2$ thermally grown—were also tested to determine if they
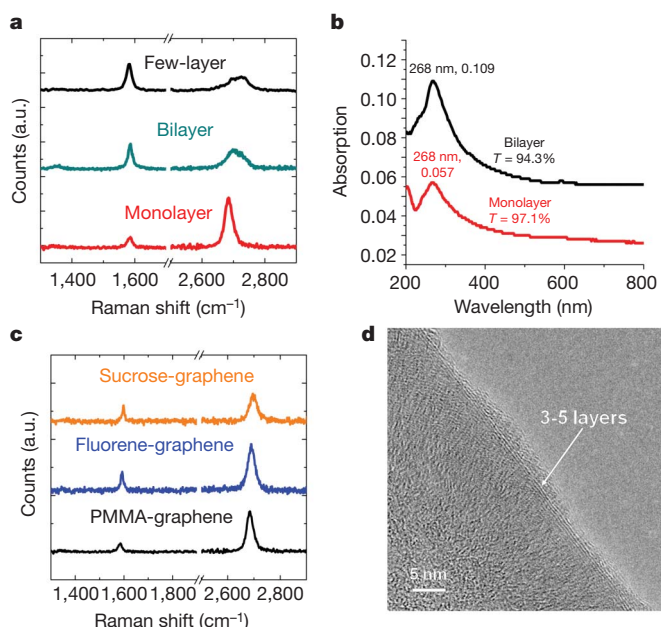
Figure 2 | Controllable growth of pristine PMMA-derived graphene films. **a**, Difference in Raman spectra from PMMA-derived graphene samples with controllable thicknesses derived from different flow rates of H_2. **b**, The ultraviolet–visible absorption spectra of monolayer graphene and bilayer graphene; peaks are labelled with wavelength of maximum absorption, and value of maximum absorption. The UV transmittance (T in %) is measured at 550 nm. **c**, Raman spectra of graphene derived from sucrose, fluorene and PMMA. **d**, HRTEM picture of PMMA-derived graphene grown on a Ni film. The PMMA-derived graphene was 3–5 layers thick at the edges.

would grow graphene when coated with PMMA. Figure 2d is the high-resolution TEM image of PMMA-derived graphene grown on a Ni catalyst, which clearly illustrates the few-layer structure around the edges of PMMA-derived graphene. The Raman spectra (Supplementary Fig. 6) confirm that Ni is an efficient catalytic substrate that converts PMMA into highly crystalline graphene materials with no D peak around $1,350\ cm^{-1}$. Under the same growth conditions, neither graphene nor amorphous carbon was obtained on Si or SiO_2 substrates, according to the Raman spectroscopic analysis of the surface after the reaction. This demonstrates the potential to grow patterned graphene from a thin film of shaped Ni or Cu deposited directly on SiO_2/Si wafers without post lithographic treatment, as PMMA-derived graphene will not grow on the Si or SiO_2 surfaces.

Pristine graphene can show weak p-type or n-type behaviour due to physisorption of small molecules, such as H_2O or NH_3 (ref. 8). However, this chemical doping effect induced by physisorption is labile, because the small molecules can be easily desorbed under heat or vacuum. Therefore, intrinsically nitrogen-doped (N-doped) graphene is more challenging to make than pristine graphene. Intrinsically N-doped graphene has been obtained by two methods: introducing a doping gas (NH_3) into the CVD systems during the graphene growth[21] or treatment of synthesized graphene or graphene oxide with NH_3 by annealing or through plasma[20,28,29]. Here, by using solid carbon sources and solid doping reagents, doped graphene can be grown in one step without any changes to the CVD system.

A doping reagent, melamine ($C_3N_6H_6$), was mixed with PMMA and spin-coated onto the Cu surface. In order to keep the nitrogen-atom concentration in the systems, we use conditions similar to those employed for the growth of PMMA-derived graphene, except that the growth was done at atmospheric pressure (Supplementary Information). The prepared polymer films were successfully converted into N-doped graphene, with an N content of 2–3.5%. The X-ray photoemission spectroscopy (XPS) data (Fig. 3a) show the difference in C $1s$ peaks between PMMA-derived graphene and N-doped

PMMA-derived graphene. The shoulder around 287 eV can be assigned to the C–N bonding. The N $1s$ peak of N-doped PMMA-derived graphene indicates that only one type of N is present, at 399.8 eV, corresponding to quaternary N in graphene[29]. This new N $1s$ peak also has a 4 eV shift from that in melamine, which shows an N $1s$ peak at 395.8 eV (Supplementary Fig. 7). The new N $1s$ peak suggests that the N $1s$ signal does not come from the melamine, but that the N atoms are uniformly bound into the graphene structure. The D peak of this material is always present in the Raman spectra, because the heteroatoms break the graphene symmetry and thereby introduce defects that are detected by Raman analysis (Fig. 3c). The D′ peak is also found in doped graphene materials obtained by the other doping methods[22,28]. The 2D peak position and I_{2D}/I_G intensity ratio reveal that this N-doped PMMA-derived graphene is monolayer graphene. Compared to PMMA-derived graphene, the I_{2D}/I_G ratio decreased from 4 to 2, implying a successful doping, according to the previously reported electrostatically gated Raman results[30].

Doping effects were also demonstrated by FETs based on N-doped PMMA-derived graphene. The n-type behaviour shown in Fig. 3d, with the neutrality point shifted to negative gate voltage, is consistently observed for devices on the same piece of N-doped PMMA-derived graphene. After keeping these FET devices under vacuum (10^{-5}–10^{-6} torr) for 24 h, their neutrality point did not move to 0 V, indicative of the covalent bonding between carbon and nitrogen rather than just physisorption; the dopant N atoms donate free electrons to graphene. Meanwhile, the mobility of N-doped graphene calculated from the N-doped FETs was about two orders of magnitude lower than in PMMA-derived graphene[21]. Owing to the broken symmetry of the lattice structure of the N-doped graphene, the N atoms act as scattering centres that suppress its mobility[26]. Patterned hydrogenation on graphene already shows its bandgap opening[31]. Similarly, if the doping atoms are periodically dispersed in graphene's matrix, they can not only
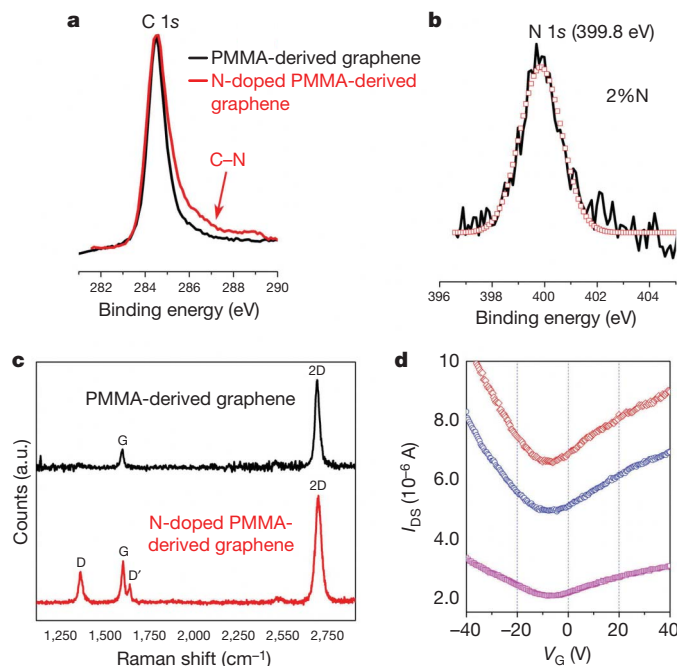
Figure 3 | Spectroscopic analysis and electrical properties of pristine and N-doped PMMA-derived graphene. **a**, XPS analysis from the C $1s$ peak of PMMA-derived graphene (black) and N-doped PMMA-derived graphene (red); the shoulder can be assigned to the C–N bond. **b**, XPS analysis, showing the N $1s$ peak (black line) and its fitting (squares), of N-doped PMMA-derived graphene. The atomic concentration of N for this sample is about 2% (C is 98%). No N $1s$ peak was observed for pristine PMMA-derived graphene. **c**, Raman spectra of pristine and N-doped PMMA-derived graphene. **d**, Room temperature I_{DS}–V_G curves ($V_{DS} = 500\ mV$) showing n-type behaviour obtained from three different N-doped graphene-based back-gated FET devices.

tune the Fermi level of graphene, but also tailor its bandgap. However, in the present N-doped graphene, the 'on'/'off' current ratio does not increase, which suggests that the N atoms are randomly incorporated into the graphene matrix. In order to manipulate both the Fermi level and the bandgap of graphene, patterned doping has to be achieved[31].

In conclusion, we have demonstrated a one-step method for the controllable growth of both pristine graphene and doped graphene using solid carbon sources. This stands as a complementary method to CVD growth while permitting growth at lower temperature.

METHODS SUMMARY

Raman spectroscopy was performed on transferred graphene films on 100 nm SiO_2/Si wafers with a Renishaw Raman microscope using 514-nm laser excitation at room temperature. The electrical properties were measured in a probe station (Desert Cryogenic TT-probe 6 system) under vacuum (10^{-5}–10^{-6} torr). The I–V data were collected by an Agilent 4155C semiconductor parameter analyser. The high-resolution TEM images were taken using a 2100F field emission gun transmission electron microscope with graphene samples directly transferred onto a C-flat TEM grid (Protochips). XPS was performed on a PHI Quantera SXM scanning X-ray microprobe with 45° takeoff angle and 100 μm beam size.

1. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. Science **306,** 666–669 (2004).
2. Geim, A. K. & Novoselov, K. S. The rise of graphene. Nature Mater. **6,** 183–191 (2007).
3. Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. Nature **438,** 197–200 (2005).
4. Ruoff, R. S. Graphene: Calling all chemists. Nature Nanotechnol. **3,** 10–11 (2008).
5. Reina, A. et al. Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. Nano Lett. **9,** 30–35 (2009).
6. Li, X. et al. Large-area synthesis of high-quality and uniform graphene films on copper foils. Science **324,** 1312–1314 (2009).
7. Kim, K. S. et al. Large-scale pattern growth of graphene films for stretchable transparent electrodes. Nature **457,** 706–710 (2009).
8. Lin, Y. et al. Operation of graphene transistors at gigahertz frequencies. Nano Lett. **9,** 422–426 (2009).
9. Lin, Y. et al. 100-GHz transistors from wafer-scale epitaxial graphene. Science **327,** 662 (2010).
10. Schedin, F. et al. Detection of individual gas molecules adsorbed on graphene. Nature Mater. **6,** 652–655 (2007).
11. Stankovich, S. et al. Graphene-based composite materials. Nature **442,** 282–286 (2006).
12. Stoller, M. D. et al. Graphene-based ultracapacitors. Nano Lett. **8,** 3498–3502 (2008).
13. Schlapbach, L. & Züttel, A. Hydrogen-storage materials for mobile applications. Nature **414,** 353–358 (2001).
14. Hernandez, Y. et al. High-yield production of graphene by liquid-phase exfoliation of graphite. Nature Nanotechnol. **3,** 563–568 (2008).
15. Stankovich, S. et al. Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphene oxide. Carbon **45,** 1558–1565 (2007).
16. Berger, C. et al. Electronic confinement and coherence in patterned epitaxial graphene. Science **312,** 1191–1196 (2006).
17. Zheng, M. et al. Metal-catalyzed crystallization of amorphous carbon to graphene. Appl. Phys. Lett. **96,** 063110 (2010).
18. Li, X. et al. Chemically derived, ultrasmooth graphene nanoribbon semiconductors. Science **319,** 1229–1232 (2008).
19. Panchakarla, L. S. et al. Synthesis, structure, and properties of boron- and nitrogen-doped graphene. Adv. Mater. **21,** 4726–4730 (2009).
20. Wang, X. et al. N-doping of graphene through electrothermal reactions with ammonia. Science **324,** 768–771 (2009).
21. Wei, D. et al. Synthesis of N-doped graphene by chemical vapor deposition and its electrical properties. Nano Lett. **9,** 1752–1758 (2009).
22. Ci, L. et al. Atomic layer of hybridized boron nitride and graphene domains. Nature Mater. **9,** 430–435 (2010).
23. Rao, C. N. R. et al. Some novel attributes of graphene. J. Phys. Chem. Lett **1,** 572–580 (2010).
24. Ferrari, A. C. et al. Raman spectrum of graphene and graphene layers. Phys. Rev. Lett. **97,** 187401–187404 (2006).
25. Chen, J. H. et al. Charge-impurity scattering in graphene. Nature Phys. **4,** 377–381 (2008).
26. Zhang, Y. et al. Direct observation of widely tunable bandgap in bilayer graphene. Nature **459,** 820–823 (2009).
27. Ci, L. et al. Controlled nanocutting of graphene. Nano Res. **1,** 116–122 (2008).
28. Lin, Y. et al. Controllable graphene N-doping with ammonia plasma. Appl. Phys. Lett. **96,** 133110 (2010).
29. Li, X. et al. Simultaneous nitrogen doping and reduction of graphene oxide. J. Am. Chem. Soc. **131,** 15939–15944 (2009).
30. Das, A. et al. Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor. Nature Nanotechnol. **3,** 210–215 (2008).
31. Balog, R. et al. Bandgap opening in graphene induced by patterned hydrogen adsorption. Nature Mater. **9,** 315–319 (2010).

Author Contributions Z.S. designed the experiments, discovered the procedures for graphene growth, performed the spectroscopic characterizations and analysis and wrote the manuscript. Z.Y. optimized the growth conditions and contributed to the spectroscopic characterizations. J.Y. performed the electrical measurements and analysis. E.B. contributed to the electrical measurements and analysis. Y.Z. carried out the sheet resistance and transmittance measurements. J.M.T oversaw all research phases and revised the manuscript. All authors discussed and commented on the manuscript.

LETTER

Impaired hydroxylation of 5–methylcytosine in myeloid cancers with mutant *TET2*

Myunggon Ko[1]*†, Yun Huang[1]*†, Anna M. Jankowska[2], Utz J. Pape[1,3], Mamta Tahiliani[1], Hozefa S. Bandukwala[1], Jungeun An[1]†, Edward D. Lamperti[1], Kian Peng Koh[1], Rebecca Ganetzky[2], X. Shirley Liu[3], L. Aravind[4], Suneet Agarwal[5], Jaroslaw P. Maciejewski[2] & Anjana Rao[1]†

TET2 is a close relative of TET1, an enzyme that converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA[1,2]. The gene encoding TET2 resides at chromosome 4q24, in a region showing recurrent microdeletions and copy-neutral loss of heterozygosity (CN-LOH) in patients with diverse myeloid malignancies[3]. Somatic *TET2* mutations are frequently observed in myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN), MDS/MPN overlap syndromes including chronic myelomonocytic leukaemia (CMML), acute myeloid leukaemias (AML) and secondary AML (sAML)[4–12]. We show here that *TET2* mutations associated with myeloid malignancies compromise catalytic activity. Bone marrow samples from patients with *TET2* mutations displayed uniformly low levels of 5hmC in genomic DNA compared to bone marrow samples from healthy controls. Moreover, small hairpin RNA (shRNA)-mediated depletion of *Tet2* in mouse haematopoietic precursors skewed their differentiation towards monocyte/macrophage lineages in culture. There was no significant difference in DNA methylation between bone marrow samples from patients with high 5hmC versus healthy controls, but samples from patients with low 5hmC showed hypomethylation relative to controls at the majority of differentially methylated CpG sites. Our results demonstrate that Tet2 is important for normal myelopoiesis, and suggest that disruption of TET2 enzymatic activity favours myeloid tumorigenesis. Measurement of 5hmC levels in myeloid malignancies may prove valuable as a diagnostic and prognostic tool, to tailor therapies and assess responses to anticancer drugs.

We transiently transfected HEK293T cells with Myc-tagged murine Tet2 and assessed 5mC and 5hmC levels by immunocytochemistry (Fig. 1 and Supplementary Figs 1–4). Myc–Tet2-expressing cells displayed a strong increase in 5hmC staining and a concomitant decrease in 5mC staining in the nucleus (Fig. 1b, c, quantified in Supplementary Fig. 4). In contrast, 5hmC was undetectable or barely detected in nuclei of cells expressing mutant Tet2 with H1302Y, D1304A substitutions in the signature HxD motif[1,12,17] involved in coordinating Fe^{2+}, and there was no obvious decrease in nuclear 5mC staining (Fig. 1b, c and Supplementary Fig. 4). These studies confirm that Tet2 is a catalytically active enzyme that converts 5mC to 5hmC in genomic DNA[13].

Mutations in TET2 residues H1881 and R1896, predicted to bind Fe^{2+} and 2-oxoglutarate (2OG), respectively, have been identified repeatedly in patients with myeloid malignancies[4,5,7,10]. HEK293T cells expressing Tet2 mutants H1802R and H1802Q (Fig. 1a and Supplementary Fig. 2) showed greatly diminished 5hmC staining and no loss of 5mC staining, consistent with participation of this residue in catalysis (Fig. 1b, c and Supplementary Fig. 4a, b). We analysed missense mutations identified in TET2 in our own (Supplementary Table 1) and other studies[3–6,11] (P1367S, W1291R, G1913D, E1318G

and I1873T). HEK293T cells expressing Tet2 mutants P1287S, W1211R or C1834D (Supplementary Figs 2 and 3a) displayed low 5hmC staining and strong 5mC staining (Supplementary Figs 3b, c and 4c, d), indicating a role for these residues in the integrity of the catalytic or DNA binding domains. Cells expressing Tet2(R1817S/M) (Fig. 1a, Supplementary Figs 2 and 3a) were positive for 5hmC staining but changes in 5mC staining could not be assessed reliably (Fig. 1b, c, Supplementary Figs 3b, c and 4).

To quantify these findings, we developed dot blot assays to detect 5hmC in genomic DNA (Supplementary Fig. 5). In the first assay format, the blot was developed with a specific antiserum to 5hmC (Supplementary Fig. 5b, left), whose ability to recognize 5hmC depended strongly on the density of 5hmC in DNA (Supplementary Fig. 5c, top). We therefore developed a more sensitive and quantitative assay in which DNA was treated with bisulphite to convert 5hmC to cytosine 5-methylenesulphonate (CMS)[14] (Supplementary Fig. 5a), after which CMS was measured with a specific anti-CMS antiserum (Supplementary Fig. 5b, right). Unlike anti-5hmC which reacted efficiently only with DNA containing high densities of 5hmC, the anti-CMS antiserum recognized DNA with an average of only a single 5hmC per 201 base pairs (Supplementary Fig. 5c, bottom). This lack of density dependence allowed us to plot the signal obtained with twofold dilutions of a standard oligonucleotide containing a known amount of 5hmC against the amount of CMS obtained after bisulphite conversion. We assumed 100% conversion efficiency[15] and used the linear portion of the standard curve to compute the amount of CMS, and therefore 5hmC, in the DNA samples (for example, see Fig. 2a, right).

To assess 5hmC levels, we obtained uniform populations of Tet2-expressing HEK293T cells by transfection with Tet2-IRES-CD25 plasmid followed by magnetic isolation of CD25-expressing cells[1]. Wild-type and mutant Tet2 proteins were expressed at comparable levels (Fig. 1d and Supplementary Fig. 3d). Anti-5hmC/CMS dot blots of genomic DNA revealed, as expected, that 5hmC was barely detectable in DNA from cells transfected with empty vector; DNA from cells expressing wild-type Tet2 showed a substantial increase in 5hmC and a corresponding decrease in 5mC; and DNA from cells expressing the HxD mutant Tet2 protein had very low 5hmC (Fig. 1e, Supplementary Figs 3e and 6). DNA from cells expressing seven of the nine mutant Tet2 proteins tested—H1802Q/R, R1817S/M, W1211R, P1287S and C1834D—contained significantly less 5hmC than DNA from cells expressing wild-type Tet2 (Fig. 1e, Supplementary Figs 3e and 6), confirming our previous conclusion that these mutations impair enzymatic activity.

We measured 5hmC (CMS) levels in genomic DNA extracted from bone marrow or blood (with >20% immature myeloid cells) of 88

[1]Department of Pathology, Harvard Medical School, Immune Disease Institute and Program in Cellular and Molecular Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [2]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, and Department of Hematologic Oncology and Blood Disorders, Cleveland Clinic, Cleveland, Ohio 44195, USA. [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. [4]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. [5]Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. †Present address: La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA (M.K., Y.H., J.A., A.R.).
*These authors contributed equally to this work.
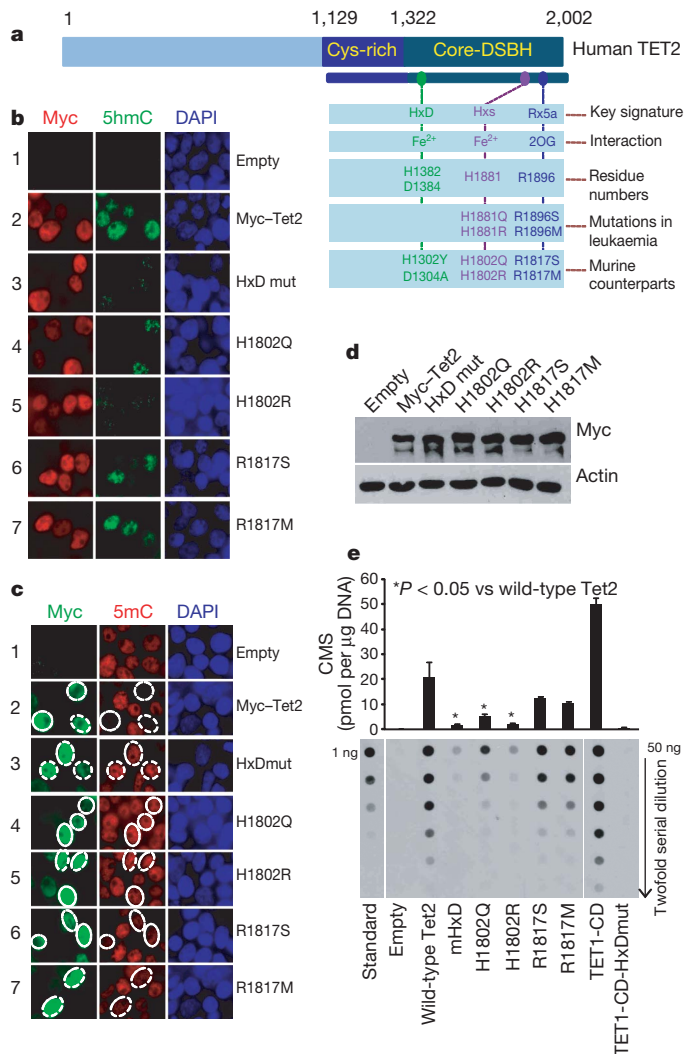
Figure 1 | The catalytic activity of Tet2 is compromised by mutations in predicted catalytic residues. a, Schematic representation of TET2. The catalytic core region contains the cysteine-rich (Cys-rich) and double-stranded beta-helix (DSBH) domains. Three signature motifs conserved among 2OG- and Fe^{2+}-dependent dioxygenases are shown[1,2]. Substitutions in the HxD signature that impair the catalytic activity of TET1 (ref. 1), leukaemia-associated mutations in the carboxy-terminal signature motifs, and corresponding substitutions introduced into murine Tet2 are indicated. **b**, Tet2 expression results in increased 5hmC by immunocytochemistry. HEK293T cells transfected with Myc-tagged wild-type and mutant Tet2 were co-stained with antibody specific for the Myc epitope (red) and antiserum against 5hmC (green). DAPI (blue) indicates nuclear staining. **c**, Tet2 expression results in loss of nuclear 5mC staining. HEK293T cells transfected with wild-type and mutant Myc-tagged Tet2 were co-stained with antibody specific for the Myc epitope (green) and antiserum against 5mC (red). **d**, Equivalent expression of wild-type and mutant Myc–Tet2. CD25[+] cells were isolated from HEK293T cells transfected with bicistronic Tet2-IRES-human CD25 plasmids, and Tet2 expression in whole cell lysates was detected by immunoblotting with anti-Myc. β-actin serves as a loading control. **e**, Genomic DNA purified from CD25[+] HEK293T cells overexpressing wild-type or mutant Tet2 was treated with bisulphite to convert 5hmC to CMS (Supplementary Fig. 5a). CMS was quantified by dot blot assay using anti-CMS and a synthetic bisulphite-treated oligonucleotide containing a known amount of CMS. As positive and negative controls, we included DNA from CD25[+] HEK293T cells transfected with TET1 catalytic domain (TET1-CD) or TET1-CD with mutations in the HxD motif (TET1-CD-HxDmut)[1].

patients with myeloid malignancies and 17 healthy controls (Supplementary Table 1). In blinded experiments, DNA was treated with bisulphite and CMS levels were measured. *TET2* mutations were strongly associated with low genomic 5hmC (Fig. 2 and Supplementary Fig. 7a). To confirm these conclusions in a statistically rigorous
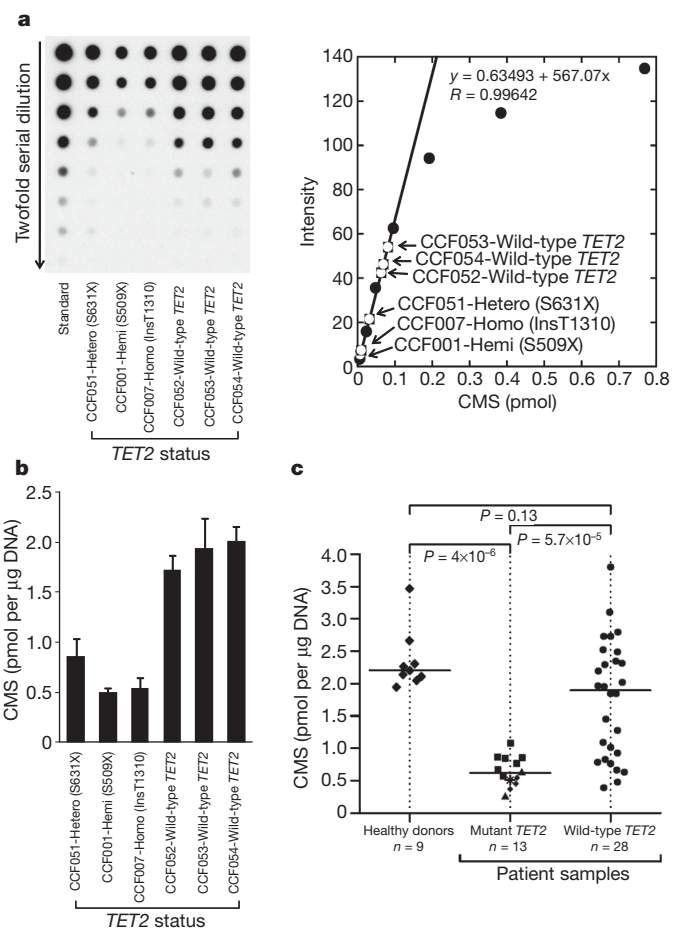
Figure 2 | TET2 mutational status correlates with 5hmC levels in patients with myeloid malignancies. a, Quantification of 5hmC by anti-CMS dot blot. Left, a representative dot blot of genomic DNA isolated from bone marrow aspirates of patients with MDS/MPN and *TET2* mutational status as indicated. A synthetic oligonucleotide with a known amount of CMS was used as standard. Right, the linear portion of the standard curve was used to estimate the amount of 5hmC in DNA from patient samples. **b**, Bar graph of data from panel **a**. The three patients with *TET2* mutations show lower 5hmC levels than the three patients with wild-type *TET2*. Error bars indicate s.d. ($n = 3$). **c**, Correlation of 5hmC levels with *TET2* mutational status. CMS levels in bone marrow samples from healthy donors and patients with myeloid malignancies (Supplementary Table 1) are shown as the median of triplicate measurements (Supplementary Fig. 7b). In the *TET2* mutant group, squares, triangles, diamonds and the star indicate homozygous, hemizygous, heterozygous and biallelic heterozygous mutations, respectively (for detailed definition, see Supplementary Methods). The horizontal bar indicates the median for each group. *P*-values for group comparisons were calculated by a two-sided Wilcoxon rank sum test. Patients bearing *TET2* mutations show uniformly low 5hmC expression levels.

fashion, we tested samples for which a sufficient amount of DNA was available to make independent dilutions in triplicate, so that a median and standard deviation for 5hmC (CMS) levels in each patient could be derived (Supplementary Fig. 7b). Analysis of DNA from 9 healthy donors and 41 patients (28 with wild-type *TET2* and 13 with *TET2* mutations, Supplementary Table 1) revealed a strong, statistically significant correlation of *TET2* mutations with low 5hmC (Fig. 2c). In contrast, samples from patients with wild-type *TET2* showed a bimodal distribution, with 5hmC levels ranging from ~0.4 to ~3.8 pmol per μg DNA (Fig. 2c, Supplementary Fig. 7, also see Fig. 4).

We examined *Tet2* expression in haematopoietic cell subsets isolated from bone marrow and thymus of C57BL/6 mice (Supplementary Figs 8 and 9). *Tet2* mRNA was highly expressed in lineage-negative (Lin[−]) Sca-1[+]c-Kit[hi] (Sca-1 is also known as Ly6a) multipotent

progenitors (LSK), at levels similar to those in embryonic stem cells (ESC). Expression was maintained at high levels in myeloid progenitors (common myeloid progenitors, CMPs, and granulocyte-monocyte progenitors, GMPs), was low in mature granulocytes (Gr-1$^+$Mac-1$^+$, also known as Ly6g and Cd11b or Itgam, respectively) and high in monocytes (Gr-1$^-$Mac-1$^+$) (Supplementary Fig. 9a, middle panel).

To test the role of Tet2 in myelopoiesis, we transduced bone marrow stem/progenitor cells with *Tet2* shRNA (Supplementary Fig. 10a), effectively downregulating *Tet2* mRNA and protein relative to control cells transduced with empty vector or scrambled shRNA (Fig. 3a, b) (refer to Supplementary Fig. 10b for choice of *Tet2* shRNA). Tet2 depletion promoted expansion of Mac-1$^+$ F4/80$^+$ (also known as Emr1) and Mac-1$^+$ CD115$^+$ (also known as Csf1r or M-CSFR, macrophage colony stimulating factor receptor) monocyte/macrophage cells in the presence of G-CSF (granulocyte colony stimulating factor) or GM-CSF (granulocyte-macrophage colony-stimulating factor), cytokines that support granulocyte and granulocyte/ monocyte development respectively, but not in the presence of M-CSF (macrophage

Figure 3 | Tet2 regulates myeloid differentiation. **a**, **b**, *Tet2* shRNA represses *Tet2* mRNA and protein expression. **a**, c-Kit$^+$ stem/progenitor cells from bone marrow of C57BL/6 mice were transduced with retroviruses (Supplementary Fig. 10). After selection with puromycin for 3 days, *Tet2* mRNA expression was assessed by quantitative RT–PCR (PCR with reverse transcription). Error bars show the range of duplicates. **b**, HEK293T cells were cotransfected with expression plasmids encoding Myc-tagged Tet2 and retroviral shRNAs. Tet2 protein expression was quantified 48 h later by anti-Myc immunoblotting of whole-cell extracts. **c**, Effect of Tet2 depletion on myeloid differentiation. Lin$^-$ cells purified from bone marrow of C57BL/6 mice were transduced with control (scramble) or shTet2 retroviruses, then grown in the presence of 50 ng ml^{-1} stem cell factor (SCF), puromycin (2 μg ml^{-1}) and cytokines (10 ng ml^{-1}) as indicated (also see Supplementary Fig. 10). After 4 days, flow cytometric analysis of Mac-1 versus F4/80 (left panel) or CD115 (right panel) was performed. All cells were GFP$^+$ on the day of analysis.
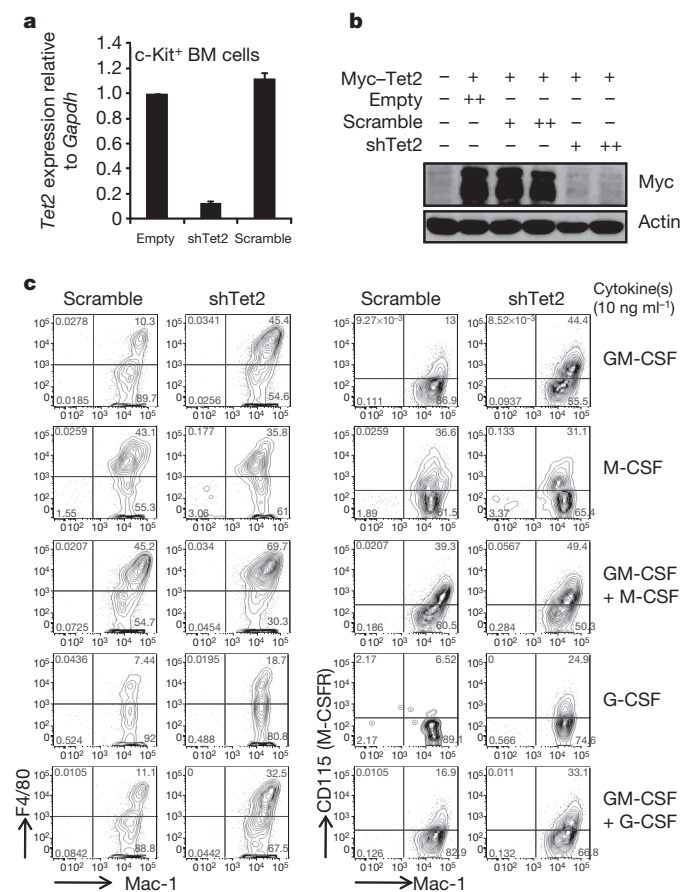
colony-stimulating factor), which promotes growth of monocytic progenitors (Fig. 3c and Supplementary Fig. 10d). Simultaneous treatment with GM-CSF and M-CSF, or GM-CSF and G-CSF, also led to increased numbers of monocyte/macrophage cells (Fig. 3c). These results indicate that Tet2 has an important role in normal myelopoiesis. However, Tet2 does not markedly influence short-term proliferation of myeloid-lineage cells: when shRNA-transduced Lin$^-$ cells were cultured in the presence of GM-CSF and pulse-labelled with bromodeoxyuridine (BrdU), Tet2 depletion promoted monocyte/ macrophage expansion but CD115$^+$ (M-CSFR$^+$) cells from the two cultures showed no difference in acute BrdU incorporation (Supplementary Fig. 11).

We asked whether 5hmC levels in tumour samples correlated with DNA methylation status. A histogram of normalized values from 88 patients and 17 healthy individuals showed the expected bimodal distribution (see Supplementary Methods): healthy controls and most patient samples with wild-type *TET2* had high 5hmC, whereas the majority of patient samples with mutant *TET2* had low 5hmC (Fig. 4b). The DNA methylation status of 62 samples was interrogated at 27,578 CpG sites. As expected[16], the resulting histograms were strikingly bimodal, with sites within and outside CpG islands showing low and predominantly high methylation, respectively (Fig. 4c). Comparison of 28 control samples with 24 high 5hmC tumour samples (22 wild-type *TET2*, 2 mutant *TET2*) showed no significant difference in DNA methylation; in contrast, comparison of the control samples with 29 low 5hmC tumour samples (7 wild-type *TET2*, 22 mutant *TET2*) yielded 2,512 differentially methylated sites, of which the majority (2,510 sites) were hypomethylated compared to controls (Fig. 4d and Supplementary Table 2). Thus TET2 loss-of-function is predominantly associated with decreased methylation at CpG sites.
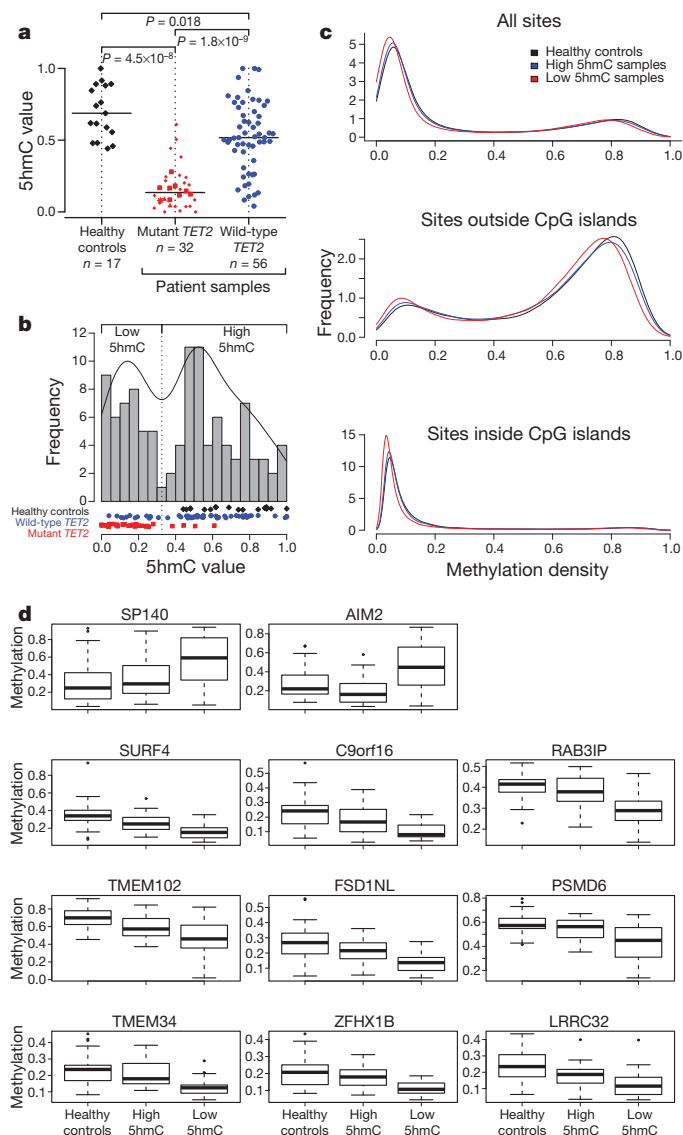
To summarize, our studies demonstrate a strong correlation between myeloid malignancies and loss of TET2 catalytic activity. The leukaemia-associated missense mutations associated with diminished 5hmC levels provide clues to the structure of the TET2 catalytic domain. The W1211R, P1287S and C1834D mutations affect positions that are highly conserved within the catalytic domain of the TET subfamily of dioxygenases[2]: W1211 is located at the beginning of the strand just amino-terminal to the core of the double-stranded beta-helix (DSBH), and is predicted to constitute part of the 'mouth' of the active site pocket of the enzyme; P1287 is predicted to stabilize the conformation of the junction between the N-terminal helix and the first core strand of the DSBH; and G1913/C1834 is predicted to be the N-terminal capping residue of a helix that lines the 'mouth' of the DSBH and potentially interacts with substrate DNA[2]. The E1238G mutation had no detectable effect on 5hmC production in our overexpression assays; however, the patient with this mutation also showed CN-LOH spanning 4q24, a feature that likely contributes to the significant reduction in 5hmC levels observed in the bone marrow.

Low 5hmC levels were observed in a subset of patients with apparently wild-type *TET2*, whose clinical phenotypes resembled those of patients with mutant *TET2*. In several of these patients, *TET2* mRNA expression was not significantly different from controls; mutations in other TET proteins have not been described (Supplementary Text). Some patients in the wild-type *TET2*/low 5hmC category may harbour mutations in regulatory or partner proteins for TET2, or in *cis*-regulatory regions controlling *TET2* mRNA expression. Alternatively, the primary event in some of these patients may be CpG hypomethylation, resulting in decreased 5hmC secondary to depletion of the substrate, 5mC.

There is little consensus on whether *TET2* mutations correlate with clinical outcome. One study reported an association with decreased survival in AML[4], whereas others report little prognostic value in MPN diseases[7,10,12]. Assays for 5hmC may increase our options for the molecular classification of myeloid malignancies, making it possible to ask whether patients with high or low levels of genomic 5hmC show differences in disease progression or therapeutic response. Notably, histone deacetylase and DNA methyltransferase inhibitors show clinical efficacy

Figure 4 | Relation of 5hmC levels to DNA methylation status.
a, Normalized 5hmC (CMS) levels in DNA from three different groups: healthy controls (black diamonds), patients with mutant *TET2* (red symbols) and patients with wild-type *TET2* (blue circles). Among *TET2* mutants, we distinguish homozygous (squares), hemizygous (triangles), heterozygous (small diamonds) and biallelic heterozygous (star) mutations (for definitions see Supplementary Methods). The horizontal bar indicates the median for each group. The number of samples in each group is indicated. **b**, Histogram of normalized 5hmC (CMS) levels in DNA from healthy donors (black diamonds), patients with mutant *TET2* (red rectangles) and patients with wild-type *TET2* (blue circles). The frequency was calculated based on a Gaussian kernel estimator. The local minimum between both modes was used as a threshold (vertical dotted line) between low and high 5hmC values. **c**, Density of methylation values for healthy controls (black), high 5hmC samples (blue) and low 5hmC samples (red) of all sites (top panel), sites outside CpG islands (middle panel) and sites inside CpG islands (lower panel). **d**, Box plot for group-specific methylation for the only two hypermethylated sites (*SP140*, *AIM2*; top panel) and the top nine hypomethylated sites (lower panels) between healthy controls and low 5hmC samples (total number of differentially methylated sites was 2,512).

in patients with CMML and AML[17]; and genomic 5hmC levels could potentially be a useful prognostic indicator or predictor of patient responses or refractoriness to 'epigenetic' therapy with demethylating agents.

DNA methylation is highly aberrant in cancer[18–20]. Because TET operates on 5mC, we were surprised to find that TET2 loss-of-function

in myeloid tumours was associated with widespread hypomethylation rather than the expected hypermethylation at differentially-methylated CpG sites. Tumour samples with low 5hmC may have expanded cells with localized hypomethylation at these sites, or TET2 may control DNA methylation indirectly, for instance by regulating the expression or recruitment of one or more DNA methyltransferases, perhaps via 5hmC-binding proteins. Alternatively, if TET2 and 5hmC are required for cells to exit the stem cell state, loss of TET2 function in myeloid neoplasms may reactivate a stem-like state characterized by generalized hypomethylation and consequent genomic instability[21,22]. Indeed, hypomorphic *DNMT1* mutations associated with genome-wide DNA hypomethylation skew haematopoietic differentiation towards myelo-erythroid lineages[23], and promote the development of aggressive T-cell lymphomas due to activation and insertion of endogenous retro-viruses[24,25]. Further studies of the role of TET2 in haematopoietic differentiation should uncover the relation between TET2 loss-of-function, DNA methylation changes and myeloid neoplasia.

METHODS SUMMARY

Patient samples. Genomic DNA was extracted from bone marrow/ peripheral blood samples from healthy donors and patients with MDS, MDS/MPN, primary and secondary AMLs. Clinical features and other detailed information pertaining to the patient samples are summarized in Supplementary Table 1.

Quantitative analysis of 5hmC and CMS levels using dot blot. For CMS detection, genomic DNA was treated with sodium bisulfite using the EpiTect Bisulfite kit (Qiagen). DNA samples were denatured and twofold serial dilutions were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-Rad). The blotted membrane was washed, air-dried, vacuum-baked, blocked and incubated with anti-5hmC or anti-CMS antibody (1:1,000) and horseradish peroxidase-conjugated anti-rabbit IgG secondary antibody. To ensure equal spotting of total DNA on the membrane, the same blot was stained with 0.02% methylene blue in 0.3 M sodium acetate (pH 5.2). To compare results obtained in different experiments, we used the normalization procedure described in Supplementary Methods (see Fig. 4a, b, which incorporate data from Fig. 2 and Supplementary Fig. 6).

Methylation analysis. The DNA methylation status of bisulphite-treated genomic DNA was probed at 27,578 CpG dinucleotides using the Illumina Infinium 27k array (Illumina)[26]. Methylation status was calculated from the ratio of methylation-specific and demethylation-specific fluorophores (β-value) using BeadStudio Methylation Module (Illumina). We removed sites on the Y and X chromosomes from the analysis because of inconsistent methylation status with respect to gender (a known problem based on communication with Illumina). Calculations are based on β values, which correspond to the methylation status of a site ranging from 0 to 1, returned by Illumina's BeadStudio software. We tested sites for differential methylation using an empirical Bayes approach employing a modified t-test (LIMMA). The false discovery rate (FDR) is controlled at a level of 5% by the Benjamini–Hochberg correction.

1. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
2. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–1710 (2009).
3. Viguié, F. *et al.* Common 4q24 deletion in four cases of hematopoietic malignancy: early stem cell involvement? *Leukemia* **19,** 1411–1415 (2005).
4. Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114,** 144–147 (2009).
5. Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360,** 2289–2301 (2009).
6. Jankowska, A. M. *et al.* Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood* **113,** 6403–6410 (2009).
7. Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genet.* **41,** 838–842 (2009).
8. Levine, R. L. & Carroll, M. A common genetic mechanism in malignant bone marrow diseases. *N. Engl. J. Med.* **360,** 2355–2357 (2009).
9. Mullighan, C. G. TET2 mutations in myelodysplasia and myeloid malignancies. *Nature Genet.* **41,** 766–767 (2009).
10. Tefferi, A. *et al.* Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. *Leukemia* **23,** 900–904 (2009).
11. Tefferi, A. *et al.* Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia* **23,** 1343–1345 (2009).

12. Tefferi, A. *et al.* *TET2* mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia* **23,** 905–911 (2009).
13. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).
14. Hayatsu, H. & Shiragami, M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **18,** 632–637 (1979).
15. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5,** e8888 (2010).
16. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
17. Tefferi, A. Epigenetic alterations and anti-epigenetic therapy in myelofibrosis. *Leuk. Lymphoma* **49,** 2231–2232 (2008).
18. Smith, L. T., Otterson, G. A. & Plass, C. Unraveling the epigenetic code of cancer for therapy. *Trends Genet.* **23,** 449–456 (2007).
19. Esteller, M. Epigenetics in cancer. *N. Engl. J. Med.* **358,** 1148–1159 (2008).
20. Gal-Yam, E. N., Saito, Y., Egger, G. & Jones, P. A. Cancer epigenetics: modifications, screening, and therapy. *Annu. Rev. Med.* **59,** 267–280 (2008).
21. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1,** 239–259 (2009).
22. Lengauer, C. Cancer. An unstable liaison. *Science* **300,** 442–443 (2003).
23. Bröske, A.-M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genet.* **41,** 1207–1215 (2009).
24. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492 (2003).
25. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20,** 116–117 (1998).
26. Jiang, Y. *et al.* Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood* **113,** 1315–1325 (2009).

Author Contributions M.K. analysed the biochemical effects of patient-associated TET2 mutations and performed the *in vitro* differentiation studies; Y.H. generated and characterized the anti-CMS antiserum, developed the quantitative dot-blot assay and quantified 5hmC in DNA samples from patients and healthy controls. A.M.J., R.G. and J.P.M. provided patient and control DNA for 5hmC quantification, performed DNA methylation arrays and analysed TET2 mutational status in patients. U.J.P. and X.S.L. carried out the statistical analysis of 5hmC levels and methylation data; M.T., H.S.B. and K.P.K. provided critical reagents; J.A. and E.D.L. contributed to molecular cloning and mouse maintenance respectively; and L.A. and S.A. provided essential intellectual input. A.R. set overall goals, coordinated collaborations and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R. (arao@idi.harvard.edu and arao@liai.org) or J.P.M. (maciejj@ccf.org).

Direct conversion of human fibroblasts to multilineage blood progenitors

Eva Szabo[1], Shravanti Rampalli[1], Ruth M. Risueño[1], Angelique Schnerch[1,2], Ryan Mitchell[1,2], Aline Fiebig-Comyn[1], Marilyne Levadoux-Martin[1] & Mickie Bhatia[1,2]

As is the case for embryo-derived stem cells, application of reprogrammed human induced pluripotent stem cells is limited by our understanding of lineage specification. Here we demonstrate the ability to generate progenitors and mature cells of the haematopoietic fate directly from human dermal fibroblasts without establishing pluripotency. Ectopic expression of OCT4 (also called POU5F1)-activated haematopoietic transcription factors, together with specific cytokine treatment, allowed generation of cells expressing the pan-leukocyte marker CD45. These unique fibroblast-derived cells gave rise to granulocytic, monocytic, megakaryocytic and erythroid lineages, and demonstrated *in vivo* engraftment capacity. We note that adult haematopoietic programs are activated, consistent with bypassing the pluripotent state to generate blood fate: this is distinct from haematopoiesis involving pluripotent stem cells, where embryonic programs are activated. These findings demonstrate restoration of multipotency from human fibroblasts, and suggest an alternative approach to cellular reprogramming for autologous cell-replacement therapies that avoids complications associated with the use of human pluripotent stem cells.

Mechanisms that govern induced pluripotent stem cell (iPSC) reprogramming from human fibroblasts remain poorly understood[1]. The process is further complicated by cellular intermediates that fail to establish a stable pluripotent state, potentially due to the inability to establish the ideal expression context of reprogramming factors to complete pluripotency induction[2–5]. These intermediates co-express genes associated with several differentiated lineages (neurons, epidermis and mesoderm)[4,5], raising the possibility that under unique conditions, fibroblasts could be induced to differentiate towards specified lineages. This maybe akin to recent demonstrations where fibroblasts were converted into single cell types, such as neurons, cardiomyocytes and macrophage-like cells[6–8]. While these studies have examined fibroblast conversion in the murine model, a similar process remains to be extrapolated towards human applications.

Our preliminary observations indicated that human dermal fibroblasts (Fibs) predominantly expressing OCT4 during the pluripotent reprogramming process express lineage differentiation markers that include the human pan-haematopoietic marker CD45. While both OCT2 (also called POU2F2) and OCT1 (also called POU2F1) bind similar DNA target motifs to OCT4 (ref. 9), and play a role in lymphoid development[10–12], OCT4 is yet to be implicated in haematopoiesis. Here, by ectopic expression of the POU protein OCT4, we demonstrate and characterize direct haematopoietic fate conversion to multipotent blood progenitors from fibroblasts in the human.

Emergence of CD45+ cells from Fibs

Reprogramming towards pluripotency requires a cascade of events that encompasses generation of various intermediate cells among a rare subset of stable iPSCs capable of teratoma formation[13,14] (Supplementary Fig. 1a–c). A portion of these intermediates form colonies that possess round cellular morphology resembling haematopoietic cells (Supplementary Fig. 2a), and express the human pan-haematopoietic marker CD45 (CD45+), but lack pluripotency marker Tra-1-60 (ref. 2) that is indicative of iPSCs (Supplementary Fig. 2b, c). These Fib-derived

CD45+ cells preferentially express OCT4 while demonstrating low levels of SOX2 and NANOG (Supplementary Fig. 2d, e). This suggested that Fib-derived intermediates could acquire a distinct lineage phenotype.

On the basis of these results, we compared the role of OCT4 during colony emergence from two sources of Fibs (adult dermal and neonatal foreskin) with that of NANOG or SOX2 alone (Fig. 1a). Transduced versus untransduced Fibs were examined between 14 and 21 days post-transduction (D14–D21; Supplementary Fig. 3). Unlike untransduced Fibs, or Fibs transduced with SOX2 (Fibs^SOX2) or NANOG (Fibs^NANOG), Fibs expressing OCT4 (Fib^OCT4) gave rise to colonies (Fig. 1a, Supplementary Fig. 3b) and exhibited OCT4 expression at levels similar to those detected in established iPSCs (Fig. 1b). Fibs^OCT4 exclusively gave rise to haematopoietic-like CD45+ cells (Fig. 1c). Furthermore, CD45+ cells (CD45+Fibs^OCT4) showed an increase in OCT4 expression (Supplementary Fig. 3c) with a concomitant decrease in the fibroblast specific gene expression[15] (Fig. 1d). Approximately 1,000 genes were downregulated and an equal number upregulated at D4, resulting in a shift towards the FibCD45+ phenotype (Supplementary Table 1). To characterize and enhance emergence of CD45+ Fibs, we used Flt3 (FMS-like tyrosine kinase 3) ligand and SCF (stem cell factor), representing inductive growth factors essential for early haematopoiesis[16,17]. Treatment with FLT3LG and SCF increased the frequency of CD45+ colony emergence from Fibs^OCT4 by 4 to 6-fold, compared with untreated Fibs^OCT4 (Fig. 1e, f), while no effect was detected from control Fibs (Fig. 1e, f, Supplementary Fig. 4). These data indicate that OCT4 is sufficient to initiate emergence of CD45+ cells from multiple sources of Fibs that are responsive to stimulation by early haematopoietic growth factors.

Conversion to CD45+ does not require iPSCs

Ectopic expression of OCT4 alone has been shown to result in pluripotent reprogramming of neural progenitors[18]. Accordingly, we examined the expression of a panel of genes known to be essential
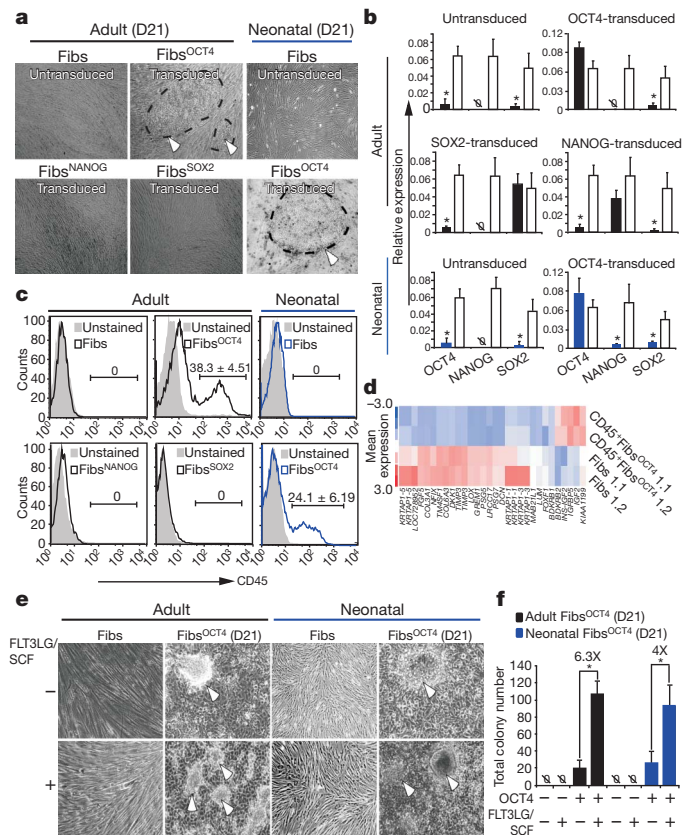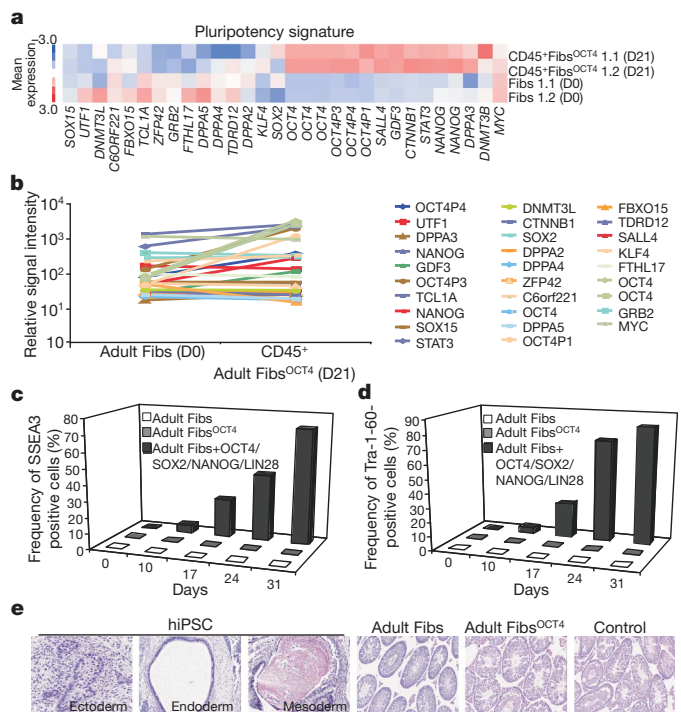
Figure 1 | OCT4 transduced human fibroblasts give rise to CD45[+ve] colonies. a, Bright field images of adult and neonatal untransduced (Fibs) and OCT4- (Fibs[OCT4]), SOX2- (Fibs[SOX2]) or NANOG- (Fibs[NANOG]) transduced Fibs at D21 (colonies, dashed lines and arrows) (n = 6). b, Relative gene expression at D21 of SOX2, NANOG and OCT4 in untransduced and transduced Fibs (shaded bars) in comparison with iPSCs (open bars; n = 3, *P < 0.001). c, FACS analysis of CD45 (pan-haematopoietic marker) levels at D21 in transduced and untransduced adult and neonatal Fibs (n = 6). d, Global gene analysis of fibroblast marker expression of Fibs at D0 versus D21. e, f, Bright field images (e) and enumeration of colonies (f) in adult and neonatal Fibs and Fibs[OCT4] with and without Flt3 and SCF (six biological replicates; colonies, white arrows; error bars, s.e.m.; *P < 0.001).

for induction and maintenance of pluripotency[14] during emergence of CD45+ Fibs. Apart from upregulation of OCT4 (Supplementary Fig. 5a), OCT4 transduction did not alter the pluripotency gene expression profile (Fig. 2a, b). Furthermore, related POU family members OCT2 and OCT1 remained unaffected (Supplementary Fig. 5b). On ectopic expression of OCT4 alone, pluripotency markers SSEA3 or Tra-1-60 were not detectable from Fibs[OCT4], whereas both gradually increased during establishment of iPSCs (Fig. 2c, d, Supplementary Fig. 5c–e). Unlike the fully reprogrammed hiPSCs (n = 8), injection of an equal number of OCT4-transduced or untransduced Fibs (n = 6) into immunodeficient mice failed to give rise to teratomas containing all three germ layers (Fig. 2e). Moreover, neither Fibs nor CD45+Fibs[OCT4] were immortalized, but could be maintained for approximately seven passages (Supplementary Fig. 6a), without elevation of MYC oncogene[19] (Supplementary Fig. 6b). Accordingly, our results indicate that the Fib[OCT4] cells manifest a haematopoietic cell fate without the detectable phenotype or functional properties of transformed or pluripotent cells.

Haematopoietic progenitor of CD45+Fibs[OCT4]

Global gene expression analysis indicated that the CD45+Fibs[OCT4] cluster with mononuclear cells derived from mobilized peripheral blood (MPB)- and umbilical cord blood (UCB)-derived haematopoietic progenitors (CD34+ cells) (Supplementary Fig. 7a, b), suggesting that


Figure 2 | OCT4 transduced dermal fibroblasts bypass the pluripotent state. a, Pluripotency gene signature at D0 versus D21. b, Pluripotency marker levels over D0 to D21. c, d, Quantitative analysis of SSEA3 (c) and Tra-1-60 (d) in Fibs, Fibs[OCT4] with and without SCF and FLT3LG, and Fibs transduced with OCT4, SOX2, NANOG, LIN28, over the human iPSC derivation timeline (n = 3). e, Teratomas derived from human iPSCs and testicular sections from mice injected with Fibs and Fibs[OCT4] (Control, saline injected).

CD45+Fibs[OCT4] may possess functional haematopoietic potential of multiple blood cell types. To functionally characterize haematopoietic capacity, both adult and neonatal CD45+Fibs[OCT4] were physically isolated and cultured with a cytokine cocktail known to support human haematopoietic progenitor development (Fig. 3a) and expansion (Supplementary Fig. 7c, d). The resulting progeny retained CD45 expression and acquired myeloid-specific markers CD33 and CD13 (Fig. 3b and Supplementary Fig. 8). A subfraction of CD45+Fibs[OCT4] progeny included monocytes expressing CD14 (Fig. 3c, d, Supplementary Fig. 9a) that could be further stimulated by M-CSF and IL-4 to mature into macrophages capable of phagocytosis[20]. CD45+Fibs[OCT4]-derived monocytes engulfed FITC-labelled latex beads (Fig. 3e, f, Supplementary Fig. 9b), unlike untransduced cytokine treated Fibs (Fig. 3e). Haematopoietic cytokine-treated CD45+Fibs[OCT4] derived from multiple sources of Fibs (adult and neonatal) also gave rise to granulocytic cell types distinct from monocytes (Supplementary Fig. 10a), as indicated by expression of CD15 (Fig. 3g, Supplementary Fig. 10b) and by characteristic cellular and polynuclear morphologies associated with neutrophils, eosinophils and basophils (Fig. 3h, Supplementary Fig. 10c). Without cytokines, CD45[+ve]Fibs[OCT4] cells retained CD45 expression, however, myeloid-specific markers were reduced and monocytic and granulocytic lineages were absent (Supplementary Fig. 11a, b). These results indicate that cytokine stimulation is necessary for haematopoietic expansion and maturation from CD45+Fibs[OCT4].

Approximately one-quarter of cytokine-stimulated CD45+Fibs[OCT4] co-expressed CD34 (Fig. 3i, Supplementary Fig. 12). In a fashion similar to somatic UCB-derived progenitors, CD45+Fibs[OCT4] gave rise to colony forming units (CFUs), indicative of the clonal proliferative developmental potential of unipotent and bipotent progenitors of the granulocytic and macrophage lineages (Fig. 3j–l). On the basis of this in vitro myeloid capacity, CD45+Fibs[OCT4] progeny were transplanted into immunodeficient NOD/SCID IL2Rγc-null (NSG) mice to characterize
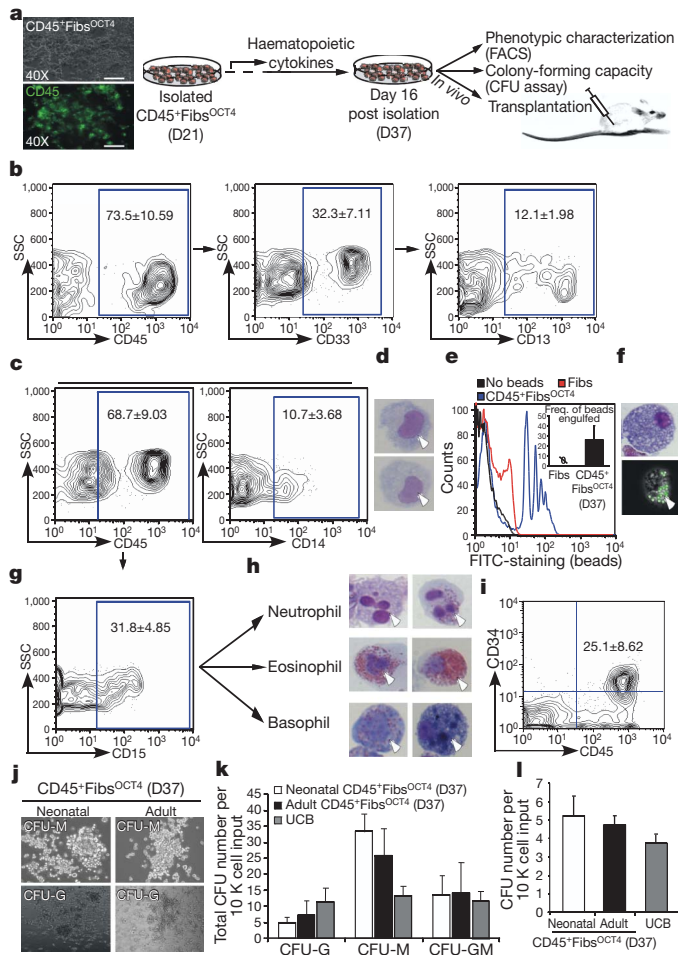
Figure 3 | *In vitro* generation of myeloid lineages from CD45$^+$ Fibs a, Schema presenting myeloid lineage derivation from CD45$^+$Fibs^{OCT4}. **b**, FACS analysis of myeloid cells derived from CD45$^+$Fibs^{OCT4} ($n = 6$). **c**, FACS analysis of monocytes and **d**, Giemsa-Wright images of monocytes (white arrow) derived from CD45$^+$Fibs^{OCT4} ($n = 6$). **e**, FACS analysis of macrophages, using FITC-labelled latex-bead uptake; inset shows quantitative analysis ($n = 3$). **f**, Giemsa-Wright stained image and immunofluorescence image of FITC-beads engulfed by macrophages (white arrow). **g**, FACS analysis of granulocytes derived from CD45$^+$Fibs^{OCT4} ($n = 6$). **h**, Giemsa-Wright stained granulocytes containing neutrophils, eosinophils and basophils (white arrows) ($n = 6$). **i**, CD45$^+$Fibs^{OCT4} derived haematopoietic progenitors ($n = 9$). **j**, Images of granulocytic (CFU-G), monocytic (CFU-M) CFUs (20×). **k**, Quantification of granulocytic, monocytic and mixed granulocytic and monocytic (CFU-GM) CFU formation from CD45$^+$CD34$^+$ cells derived from adult and neonatal Fibs and UCB ($n = 3$). **l**, CFU frequency of adult and neonatal CD45$^+$Fibs^{OCT4} and UCB derived haematopoietic progenitors ($n = 3$; *$P < 0.001$). Error bars, s.e.m.

their *in vivo* potential (Fig. 4a). CD45$^+$Fibs^{OCT4}-derived cells engrafted all transplanted NSG recipients up to levels of 20%, as indicated by HLA-A/B/C$^+$ cells (Fig. 4b), while Fibs and saline showed no engraftment (Fig. 4c). Engraftment levels of CD45$^+$Fibs^{OCT4} were comparable to UCB-derived progenitors and MPB-derived progenitors (Fig. 4d). Primary reconstituted NSG recipients exhibited a predominantly myeloid phenotype (~41% CD45$^+$CD14$^+$) (Fig. 4c), compared with UCB and MPB cells (Fig. 4e). A proportion of the engrafted cells retained CFU initiation potential similar to cells from human UCB (Fig. 4f, Supplementary Fig. 13a–d). The ability to generate haematopoietic progenitors and the presence of engraftment, albeit at low levels, in the contralateral bones of the primary NSG recipients at 10 weeks post transplant (Supplementary Fig. 14a) supports the *in vivo* functional capacity of CD45$^+$Fibs^{OCT4}-derived cells. Engrafted CD45$^+$ cells were only capable
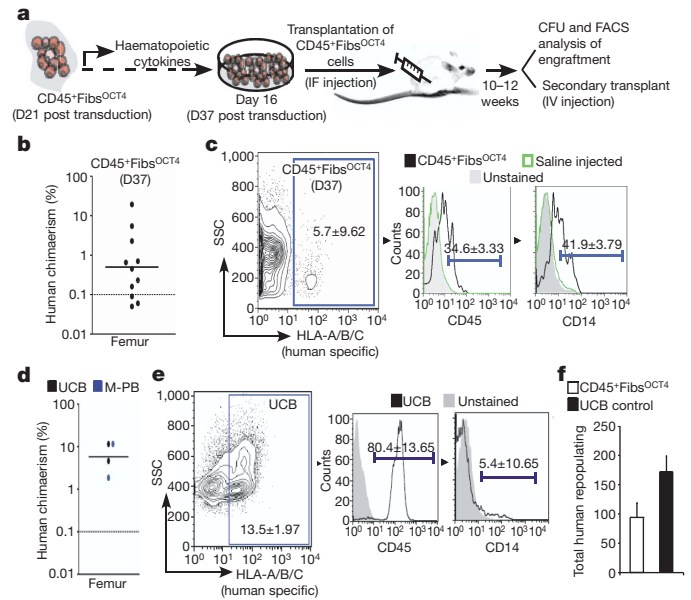

Figure 4 | *In vivo* capacity of CD45$^+$Fibs^{OCT4}. a, Schematic of NOD/SCID IL2Rγc-null mice xenograft model for injection of CD45$^+$Fibs^{OCT4}. **b**, Graph representing human chimaerism of CD45$^+$Fibs^{OCT4} treated with cytokines ($n = 12$). **c**, FACS histograms of engrafted CD45$^+$Fibs^{OCT4} showing the presence of CD45$^+$ and CD14$^+$ population ($n = 12$). **d**, Graph representing human chimaerism of UCB derived progenitors and MPB cells ($n = 4$). **e**, Representative FACS histograms of engrafted UCB and MPB cells, showing the presence of CD45$^+$ and CD14$^+$ population ($n = 4$). **f**, CFU capacity of engrafted CD45$^+$Fibs^{OCT4} versus UCB ($n = 3$).

of limited engraftment ability in secondary recipient NSG mice (Supplementary Fig. 14b), indicating that they do not possess transformed leukaemic stem cell properties[21], and thus represent safer haematopoietic transplantation product alternatives, as compared to hPSC-derived cells that retain tumour potential[22,23]. Collectively, our data indicate that CD45$^+$Fibs^{OCT4} can give rise to functional haematopoietic progenitor-like cells that are able to mature into myeloid lineages *in vitro* and *in vivo*.

Erythroid and megakaryocytic potential of CD45$^+$ Fibs

Despite the ability to derive all myeloid lineages from CD45$^+$Fibs^{OCT4}, erythroid cells were not detected. On OCT4 transduction, Fibs expressed the erythroblast marker CD71 at a frequency of nearly 40% (Fig. 5a). As Erythropoietin (EPO) induces early erythroid differentiation[24], we added EPO to Fib^{OCT4} cultures, and observed a twofold increase in the number of Fibs expressing CD71, along with an increase in expression of Glycophorin-A (Fig. 5b) and adult β-globin protein (Fig. 5c). In contrast, Fibs (Fig. 5c) and hPSC-derived haematopoietic cells (Fig. 5c inset) lacked β-globin protein. In the absence of EPO, only β-globin transcript was expressed in the CD45$^+$Fibs^{OCT4} (Fig. 5d), while β-globin protein was undetectable (Fig. 5c). In contrast, and unlike haematopoietic cells derived from hPSCs[25], haematopoietic cells derived from CD45$^+$Fibs^{OCT4} lacked embryonic globin expression and expressed modest levels of fetal globin (Fig. 5d). EPO-treated CD45$^+$Fibs^{OCT4} exhibited both primitive and mature erythrocyte (enucleated) morphologies (Fig. 5e) and allowed for erythroid progenitor emergence, detected by colony formation (BFU-E) and CFU-Mixed colonies (CFU-Mix; dual myeloid and erythroid capacity), similar to that for UCB, without reduction in monocytic or granulocytic progenitor capacity (Fig. 5f, g, Supplementary Fig. 15a, b).

Studies have indicated that erythroid and megakaryocytic lineage commitment occur together and potentially arise from a common precursor population[26,27]. EPO stimulation of CD45$^+$Fibs^{OCT4} induced emergence of megakaryocytic lineage, as detected by Megakaryocytic (Mk)-CFU assay[28], indicated by the presence of Mk-specific antigen GPIIb/IIIa (CD41; red/pink colonies) (Fig. 5h right panel, Fig. 5i), while
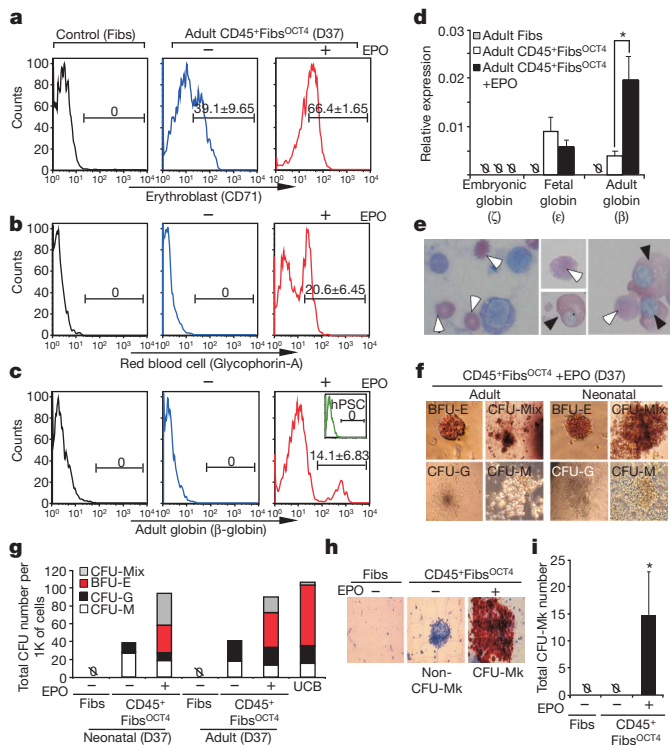
Figure 5 | EPO treated CD45$^+$Fibs^{OCT4} generate erythroid and megakaryocytic progenitors. a–c, Representative FACS histograms of erythroblast marker, CD71 (**a**), Glycophorin-A (**b**) and adult-globin (**c**; upper panel, differentiated hPSCs) in Fibs and CD45$^+$Fibs^{OCT4} with and without EPO ($n = 3$). **d,** Relative mRNA expression of embryonic, fetal and adult-globins ($n = 3$; *$P < 0.001$). **e,** Giemsa-Wright stained EPO treated CD45$^+$Fibs^{OCT4} showing primitive (black arrow) and mature (white arrow) erythrocyte morphologies. **f,** CFU images, and **g,** quantitative analysis of CFU formation of EPO treated adult and neonatal CD45$^+$Fibs^{OCT4} (20×; $n = 3$). (Erythroid blast forming units, BFU-E; all lineages, CFU-Mix). **h,** Megakaryocytic CFU (CFU-Mk) images (CD41$^+$ cells; 20×), and **i,** quantification ($n = 3$; *$P < 0.001$). ⓪, not detected.

this haematopoietic progenitor type was absent in CD45$^+$Fibs^{OCT4} without EPO (Fig. 5h middle panel, Fig. 5i) or control Fibs (Fig. 5h left panel, Fig. 5i). Taken together, EPO treatment induces definitive and not primitive (embryonic) haematopoietic programs[29], and reveals mega-karyocytic lineage during conversion of Fibs to haematopoietic fate.

OCT4 haematopoietic program from Fibs

To develop a broader understanding of the role of OCT4 during hae-matopoietic conversion of Fibs (Fig. 6a), we examined gene expression and OCT4 occupancy over the time course of CD45$^+$ cell emergence and maturation. As early as day (D) 4 post transduction, significant changes occur in molecular pathways, including metabolic and develop-mental processes (Supplementary Fig. 16a). Furthermore, global gene expression of the Fibs over the course of CD45$^+$ cell emergence (D0, D4 and D21) indicated a decrease in fibroblast-specific gene expression[15] (Fig. 6b), without pluripotency gene induction (Fig. 6c), excluding OCT4. Fibs^{OCT4} immediately upregulated haematopoietic cytokine receptors, including FLT3 and KIT, receptors of FLT3LG and SCF, respectively (Fig. 6d), along with transcription factors associated with early haematopoietic development (Fig. 6e, Supplementary Fig. 16b, c). In addition, Fibs possess low to undetectable levels of genes associated with pluripotency, such as NANOG and SOX2, or haematopoietic spe-cification, such as SCL/TAL1, RUNX1, C/EBPα, GATA1 or PU.1/SPI1 (refs 6, 30–32) (Fig. 6f, Supplementary Fig. 17a). However, transduction with OCT4 was accompanied by a substantial increase of SCL, C/EBPα, GATA1 and RUNX1 (Fig. 6f). Interestingly, PU.1 and MIXL1, which were previously shown to regulate primitive blood development[6,33,34],

were not differentially regulated (Fig. 6e, f, Supplementary Fig. 16b, c). Expression of genes associated with mesodermal transition from the pluripotent state, such as Brachyury and GATA2, were absent in both Fibs and CD45$^+$Fibs^{OCT4} (Fig. 6f), indicating that haematopoietic specification from Fibs does not involve embryonic-mesodermal tran-sitions[35,36]. Molecular analysis of cytokine treated CD45$^+$Fibs^{OCT4} (D37) revealed reduced OCT4 levels, while maintaining RUNX1, SCL and C/EBPα levels (Supplementary Fig. 17b), and expression of all adult globins, including haemoglobin-α, β and δ (Supplementary Fig. 17c, Fig. 5d).

Like OCT4, OCT1 and OCT2 are also able to regulate haematopoietic-specific genes (Supplementary Fig. 18)[37–41]. While expression of OCT4 increased during CD45$^+$ cell emergence, followed by a significant reduction on cytokine treatment, the levels of OCT2 and OCT1 remained unchanged (Fig. 6g), suggesting that OCT4 does not target other Oct family members. Nevertheless, OCT1, OCT2 and OCT4 have the potential to bind the same octamer (POU) binding sequences in a cell context specific manner, thereby raising the possibility that OCT4 has the capacity to bind and potentially regulate gene targets of OCT1 and OCT2 (refs 37–40) (Fig. 6h, Supplementary Fig. 18). Thus, we examined OCT4 occupancy of haematopoietic, non-haematopoietic and pluripotency genes that contain shared OCT1, OCT2 or OCT4 binding sequences in their putative promoters/enhancers (Fig. 6h, Supplementary Fig. 18). Consistent with changes in gene expression (Fig. 6f), RUNX1, SCL, CD45 and GATA1 displayed substantial OCT4 occupancy (Fig. 6i). To assess the specificity of OCT4 occupancy of haematopoietic targets during CD45$^+$ cell emergence, we examined non-haematopoietic associated promoters previously shown to bind OCT1 or OCT2. Consistent with global gene expression (Supplementary Fig. 7a), housekeeping genes GADD45A and POLR2A exhibited an increase in OCT4 occupancy at their respective promoters, while non-haematopoietic genes MYF5 and NKX2.5, associated with meso-dermal development, did not (Fig. 6j). In contrast to hPSCs, OCT4 did not occupy NANOG, MYC and TBX3 promoters (Fig. 6k) in CD45$^+$Fibs^{OCT4}. While OCT4 binds its own promoter (Fig. 6k), it does not bind the OCT2 promoter (Fig. 6i), consistent with OCT2 expres-sion (Supplementary Fig. 5a, b). Collectively, these temporal gene expression analyses along with OCT4 occupancy studies demonstrate that ectopic OCT4 expression orchestrates haematopoietic program potential in Fibs towards blood fate conversion.

Discussion

Our current study demonstrates the ability of human fibroblasts to be directly converted to multipotent haematopoietic progenitors of the myeloid, erythroid and megakaryocytic lineages via OCT4-dependent cellular programming without traversing the pluripotent state or activation of mesodermal pathways[35,36]. Given that transition from primitive to definitive haematopoiesis is delineated by the shift from embryonic to adult haemoglobin expression[29], our study demon-strates that unlike hPSC-derived haematopoietic cells[42], CD45$^+$Fibs exclusively acquire adult-globin protein indicative of definitive hae-matopoietic program activation. Acquisition of the haematopoietic phenotype is linked to the direct binding of OCT4 to the regulatory loci of haematopoietic-specific genes[37,40,41,43]. While OCT1 and OCT2 have been shown to play a role in adult lymphopoiesis[10–12], OCT4 has not been implicated in blood fate. Given the conservation between the native or predicted octamer binding sequences among OCT1, OCT2 and OCT4, it is highly possible that POU domains shared among OCT proteins have a redundant role in haematopoietic fate conver-sion. Although OCT4 converts fibroblasts to myeloid and erythroid progenitors, lymphoid fate was not detected. Accordingly, it is plaus-ible that ectopic expression of OCT4, OCT1 and OCT2, coupled with lymphoid specific culture conditions, may support lymphoid conver-sion from Fibs, and is currently being pursued.

Although recent reports demonstrate conversion of mouse fibro-blasts to neural, cardiac and macrophage-like cells[6–8], the present
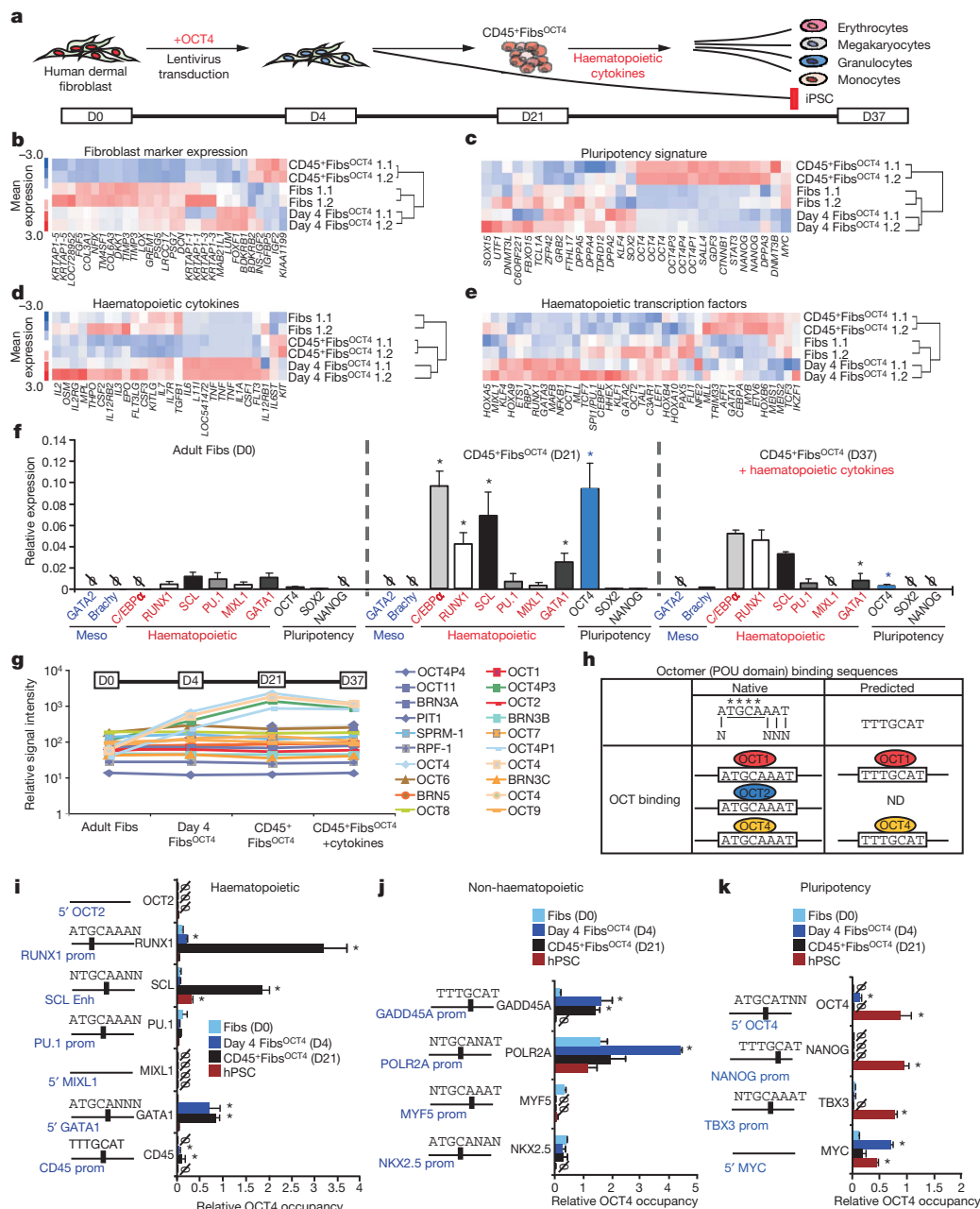
Figure 6 | OCT4 causes haematopoietic program activation in Fibs. a, Proposed model for haematopoietic fate conversion from CD45$^+$Fibs^{OCT4} (D0 to D37). **b–e**, Global gene expression based on fibroblast marker expression (**b**), pluripotency signature (**c**), haematopoietic cytokines (**d**) and haematopoietic transcription factors (**e**). **f**, Relative mRNA expression of mesodermal, haematopoietic specific and pluripotency genes at D0 versus D21 and D37 (n = 4, *P < 0.001). **g**, Gene expression profile of OCT (POU) genes.

h, Schematic of native and predicated octamer binding sequences for OCT4, OCT1 and OCT2, (N, any nucleotide; starred/underlined, conserved octamer-binding region). **i–k**, Right panels: relative OCT4 occupancy of haematopoietic specific (**i**), non-haematopoietic (**j**) and pluripotency (**k**) gene enhancer/promoter regions (n = 3; *P < 0.001). **i–k**, Left panels: proximity maps of primers (arrows) relative to native or predicted octamer-binding region (black box).

study uniquely demonstrates the ability to generate multipotent, rather than unipotent (single lineage), cell types from fibroblasts in the human, hence establishing future clinical utility of these multi-potent cells of the haematopoietic lineage. Taking into account the yield, expansion capacity and clinical feasibility[44] of using this direct conversion approach to haematopoietic fate (Supplementary Table 2), our technique could provide a reasonable basis for autologous cell replacement therapies.

METHODS SUMMARY

Cultures. Adult and neonatal dermal fibroblasts were cultured in F12-DMEM media supplemented with (1) IGFII and bFGF, or (2) IGFII, bFGF, Flt3 and SCF, on Matrigel-coated plates. Lentiviral vectors (pSIN) containing cDNAs of OCT4,

NANOG, SOX2 and LIN28 were obtained from Addgene and were transfected into 293-FT cells using the virapower packaging kit (Invitrogen). Fibroblast transductions were performed at 24 h post 10^4 seeding on Matrigel. For derivation of CD45$^+$ cells, fibroblasts were transduced with OCT4 expressing lentivirus and cultured in media (1) or (2), and iPSCs were derived as previously described[15]. Further haematopoietic differentiation was carried out using EB media supplemented with haematopoietic cytokines.

Functional/phenotype analysis. Flow cytometry analysis of haematopoietic and pluripotency markers was performed using FACSCalibur (Beckman Coulter), and analysis was performed using the FlowJo 8.8.6 software. Cell sorting was performed using FACSAria II (Becton-Dickinson); Histological profiling of hemato-poietic cells was performed using Cytospin and Giemsa-Wright staining and confirmed by the McMaster Pathology and Hematology Group; CFU formation was assayed using Methocult and Megacult kits from Stem Cell Technologies;

Macrophage phagocytosis assay was performed using Fluorescein conjugated-latex beads (Sigma) as particle tracers to analyse uptake by monocytes derived from $CD45^+Fibs^{OCT4}$ cells; *in vivo* engraftment capacity was evaluated by intrafemoral injection of $CD45^{+ve}$ cells into NSG mice. Ten weeks later bone marrow from injected femur, contralateral bones and spleen was analysed for the presence of human cells by flow cytometry; teratoma formation was evaluated by intratesticular injection into NOD/SCID mice. Resulting teratomas were evaluated for the presence of mesoderm, endoderm and ectoderm through histological examination.

Molecular analysis. For qPCR and microarray analysis, RNA was extracted using a total RNA purification kit (Norgen). Microarray analysis was done using Human Gene 1.0 ST arrays (Affymetrix) and dChIP software. OCT4 DNA occupancy (OCT4 ChIP) was done as previously described[45]. See Supplementary Methods for additional details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
2. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnol.* **27**, 1033–1037 (2009).
3. Lin, T. *et al.* A chemical platform for improved induction of human iPSCs. *Nature Methods* **6**, 805–808 (2009).
4. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
5. Kanawaty, A. & Henderson, J. Genomic analysis of induced pluripotent stem (iPS) cells: routes to reprogramming. *Bioessays* **31**, 134–138 (2009).
6. Feng, R. *et al.* PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
7. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
8. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
9. Kang, J., Shakya, A. & Tantin, D. Stem cells, stress, metabolism and cancer: a drama in two Octs. *Trends Biochem. Sci.* **34**, 491–499 (2009).
10. Brunner, C. *et al.* B cell-specific transgenic expression of Bcl2 rescues early B lymphopoiesis but not B cell responses in BOB.1/OBF.1-deficient mice. *J. Exp. Med.* **197**, 1205–1211 (2003).
11. Emslie, D. *et al.* Oct2 enhances antibody-secreting cell differentiation through regulation of IL-5 receptor α chain expression on activated B cells. *J. Exp. Med.* **205**, 409–421 (2008).
12. Pfisterer, P. *et al.* CRISP-3, a protein with homology to plant defense proteins, is expressed in mouse B cells under the control of Oct2. *Mol. Cell. Biol.* **16**, 6160–6168 (1996).
13. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
14. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
15. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
16. Hassan, H. T. & Zander, A. Stem cell factor as a survival and growth factor in human normal and malignant hematopoiesis. *Acta Haematol.* **95**, 257–262 (1996).
17. Lyman, S. D. *et al.* Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase receptor: a proliferative factor for primitive hematopoietic cells. *Cell* **75**, 1157–1167 (1993).
18. Kim, J. B. *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**, 649–653 (2009).
19. Lebofsky, R. & Walter, J. C. New Myc-anisms for DNA replication and tumorigenesis? *Cancer Cell* **12**, 102–103 (2007).
20. Silverstein, S. C., Steinman, R. M. & Cohn, Z. A. Endocytosis. *Annu. Rev. Biochem.* **46**, 669–722 (1977).
21. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nature Immunol.* **5**, 738–743 (2004).
22. Roy, N. S. *et al.* Functional engraftment of human ES cell-derived dopaminergic neurons enriched by coculture with telomerase-immortalized midbrain astrocytes. *Nature Med.* **12**, 1259–1268 (2006).
23. Amariglio, N. *et al.* Donor-derived brain tumor following neural stem cell transplantation in an ataxia telangiectasia patient. *PLoS Med.* **6**, e1000029 (2009).
24. Fried, W. Erythropoietin and erythropoiesis. *Exp. Hematol.* **37**, 1007–1015 (2009).
25. Perlingeiro, R. C., Kyba, M. & Daley, G. Q. Clonal analysis of differentiating embryonic stem cells reveals a hematopoietic progenitor with primitive erythroid and adult lymphoid-myeloid potential. *Development* **128**, 4597–4604 (2001).
26. Debili, N. *et al.* Characterization of a bipotent erythro-megakaryocytic progenitor in human bone marrow. *Blood* **88**, 1284–1296 (1996).
27. Klimchenko, O. *et al.* A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis. *Blood* **114**, 1506–1517 (2009).
28. Strodtbeck, D. *et al.* Graft clonogenicity and intensity of pre-treatment: factors affecting outcome of autologous peripheral hematopoietic cell transplantation in patients with acute myeloid leukemia in first remission. *Bone Marrow Transplant.* **36**, 1083–1088 (2005).
29. Orkin, S. H. & Zon, L. I. Hematopoiesis and stem cells: plasticity versus developmental heterogeneity. *Nature Immunol.* **3**, 323–328 (2002).
30. Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432–434 (1995).
31. Ichikawa, M., Asai, T., Chiba, S., Kurokawa, M. & Ogawa, S. Runx1/AML-1 ranks as a master regulator of adult hematopoiesis. *Cell Cycle* **3**, 722–724 (2004).
32. Friedman, A. D. Transcriptional control of granulocyte and monocyte development. *Oncogene* **26**, 6816–6828 (2007).
33. Koschmieder, S., Rosenbauer, F., Steidl, U., Owens, B. M. & Tenen, D. G. Role of transcription factors C/EBPα and PU.1 in normal hematopoiesis and leukemia. *Int. J. Hematol.* **81**, 368–377 (2005).
34. Ng, E. S. *et al.* The primitive streak gene Mixl1 is required for efficient haematopoiesis and BMP4-induced ventral mesoderm patterning in differentiating ES cells. *Development* **132**, 873–884 (2005).
35. Tsai, F. Y. *et al.* An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **371**, 221–226 (1994).
36. Vijayaragavan, K. *et al.* Noncanonical Wnt signaling orchestrates early developmental events toward hematopoietic cell fate from human embryonic stem cells. *Cell Stem Cell* **4**, 248–262 (2009).
37. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
38. Kistler, B., Pfisterer, P. & Wirth, T. Lymphoid- and myeloid-specific activity of the PU.1 promoter is determined by the combinatorial action of octamer and ets transcription factors. *Oncogene* **11**, 1095–1106 (1995).
39. Rodda, D. J. *et al.* Transcriptional regulation of nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005).
40. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364–377 (2009).
41. Ghozi, M. C., Bernstein, Y., Negreanu, V., Levanon, D. & Groner, Y. Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. *Proc. Natl Acad. Sci. USA* **93**, 1935–1940 (1996).
42. Chang, K. H. *et al.* Definitive-like erythroid cells derived from human embryonic stem cells coexpress high levels of embryonic and fetal globins with little or no adult globin. *Blood* **108**, 1515–1523 (2006).
43. Kwon, U. K., Yen, P. H., Collins, T. & Wells, R. A. Differential lineage-specific regulation of murine CD45 transcription by Oct-1 and PU.1. *Biochem. Biophys. Res. Commun.* **344**, 146–154 (2006).
44. Feugier, P. *et al.* Hematologic recovery after autologous PBPC transplantation: importance of the number of postthaw CD34+ cells. *Transfusion* **43**, 878–884 (2003).
45. Rampalli, S. *et al.* p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nature Struct. Mol. Biol.* **14**, 1150–1156 (2007).

Author Contributions All authors contributed to the acquisition, analysis and interpretation of the data; E.S., S.R., R.M.R. and M.B. initiated and designed the study; A.S. performed Affymetrix analyses; R.M.R. performed *in vivo* analyses; E.S., S.R., R.M.R. and M.B. wrote the paper.

Author Information Data are available on the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE24621. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.B. (mbhatia@mcmaster.ca).

METHODS

Cell culture. Primary human dermal adult fibroblasts derived from breast dermal tissue, and neonatal fibroblasts derived from foreskin tissue, were initially maintained in fibroblast medium (DMEM (Gibco)) supplemented with 10% v/v FBS (Neonatal Bovine Serum, HyClone), 1 mM L-glutamine (Gibco), and 1% v/v nonessential amino acids (NEAA; Gibco) before transduction with OCT4 lentivirus vector. Human dermal fibroblasts transduced with OCT4 were maintained on matrigel-coated dishes in complete F12 medium (F12 DMEM; Gibco) supplemented with 10% knockout serum replacement (Gibco), 1% NEAA (Gibco), 1 mM L-glutamine (Gibco), and 0.1 mM β-mercaptoethanol containing 16 ng ml^{-1} bFGF (BD Biosciences) and 30 ng ml^{-1} IGFII (Millipore), or in complete F12 medium containing 16 ng ml^{-1} bFGF and 30 ng ml^{-1} IGFII and supplemented with 300 ng ml^{-1} Flt-3 (R&D Systems) and 300 ng ml^{-1} stem cell factor (SCF; R&D Systems), for 21 days. The arising CD45$^+$ OCT-transduced cells were transferred onto low attachment 24-well plates in haematopoietic medium, consisting of 80% knockout DMEM (KO-DMEM) (Gibco), 20% v/v non-heat inactivated fetal calf serum (FCS) (HyClone), 1% v/v NEAA, 1 mM L-glutamine and 0.1 mM β-mercaptoethanol (Sigma), for 16 days. Cultures were replaced with haematopoietic differentiation medium with cytokines (SCF, G-CSF, FLT3LG, IL-3, IL-6 and BMP-4; R&D Systems), or for erythroid/megakaryocytic differentiation the medium was supplemented with haematopoietic cytokines plus 3 U ml^{-1} EPO and changed every 4 days. This was followed by collection for molecular and functional analysis.

Lentivirus production. Lentiviral vectors (pSIN) containing cDNAs of OCT4, NANOG, SOX2 and Lin-28 were obtained from Addgene. These vectors were transfected with virapower packaging kit from Invitrogen in a 293-FT packaging cells line. Viral supernatants were harvested 48 h after transfection and ultracentrifuged to concentrate the virus. An equal amount of each virus was used for fibroblast transduction in the presence of 8 µg ml^{-1} polybrene.

Lentivirus transduction. For generation of cells containing single transcription factors, human adult dermal fibroblasts (Fibs) (derived from breast skin; age 30–40 yr) or neonatal foreskin Fibs were seeded at a density of 10,000 cells per well on matrigel coated 12-well plates. Twenty-four hours after seeding, Fibs were infected with lentivirus expressing either OCT4 or NANOG or SOX2 (NANOG and SOX2 transduction was only performed for adult dermal Fibs). Transduced fibroblasts were then grown in complete F12 medium containing 16 ng ml^{-1} bFGF and 30 ng ml^{-1} IGFII supplemented with 300 ng ml^{-1} Flt-3 and 300 ng ml^{-1} SCF, or complete F12 medium containing 16 ng ml^{-1} bFGF and 30 ng ml^{-1} IGFII alone, for up to 21 days. Emerging CD45$^+$ colonies were counted 14–21 days after infection. Colonies were picked manually and maintained on matrigel-coated wells. Molecular analysis was done on purified untransduced Fibs (D0), OCT4 transduced Fibs at day 4 (D4), CD45$^+$ Fibs at day 21 (D21) and haematopoietic cytokine treated or untreated CD45$^+$ Fibs at day 37 (D37). Day 4 post OCT4 transduction was chosen as the early event time point, based on a number of criteria: optimal time for recovery following transduction; visible morphological changes within the culture; and resumption of normal cell cycle kinetics. The day 4 OCT4 transduced Fibs (D4) were isolated by puromycin selection overnight (OCT4 vector contains a puromycin resistance cassette), and the purity of the sample was validated by staining for OCT4 followed by OCT4 expression analysis using flow cytometry; samples used for molecular analysis exhibited 99% OCT4 levels. The D21 and D37 CD45$^+$ Fibs^{OCT4} were isolated on the basis of their CD45 expression. D21 and D37 cells were stained with CD45-APC antibody (BD Biosciences) and sorted using a FACSAria II (Becton Dickinson); samples used for molecular analysis exhibited 99% CD45 levels.

Induction of reprogramming. For generation of reprogrammed cells from fibroblasts, cells were seeded at the density of 10,000 cells per well on matrigel coated 12-well plates. Twenty-four hours after seeding, fibroblasts were transduced with lentivirus expressing OCT4/NANOG/SOX2/LIN28 (ref. 15). Transduced fibroblasts were then grown in F12 medium supplemented with 30 ng ml^{-1} IGFII and 16 ng ml^{-1} bFGF. Reprogrammed iPSC colonies were counted four weeks post infections. Colonies were picked manually and maintained on matrigel-coated wells.

Live staining. For live staining, sterile Tra-1-60 antibody (Millipore) was pre-conjugated with sterile Alexa Fluor-647 at room temperature. Reprogrammed colonies were washed once with F12 medium and incubated with Tra-1-60-Alexa 647 antibodies for 30 min at room temperature. Cultures were then washed twice to remove unbound antibody. Cells were visualized by an Olympus IX81 fluorescence microscope.

Flow cytometry. For pluripotency marker expression, cells were treated with collagenase IV, and then placed in cell dissociation buffer for 10 min at 37 °C (Gibco). Cell suspensions were stained with SSEA3 antibody (1:100) (Developmental Studies Hybridoma Bank, mAB clone MC-631, Univ. Iowa) or Tra-1-60-PE (1:100) antibody (BD Biosciences). For SSEA3 staining, Alexa Fluor-647 goat anti-rat IgM (1:1,000) (Molecular Probes, Invitrogen) was used as the secondary antibody.

Live cells were identified by 7-amino actinomycin (7AAD) exclusion and then analysed for cell surface marker expression using a FACSCalibur (Becton Dickinson). Collected events were analysed using FlowJo 8.8.6 Software (Tree Star).

Cells from the haematopoietic differentiation medium were disassociated with TrypLE (Gibco) at D16 and analysed for expression of haematopoietic progenitor and mature haematopoietic markers. Haematopoietic cells were identified by staining single cells with fluorochrome-conjugated monoclonal antibodies (mAb): CD34-FITC and APC- or FITC-labelled anti-human CD45 (BD Biosciences), FITC-anti-CD33 (BD Pharmingen), PE-anti-CD13 (BD Pharmingen), PE- or FITC-anti-CD71 (BD Pharmingen), FITC-anti-HLA-A/B/C (BD Pharmingen), PE-anti-CD15 (BD Pharmingen), PE-anti-CD15 (BD Pharmingen); PE anti-CD14 (BD Pharmingen), FITC- or PE-anti-GlyA (BD Pharmingen), and APC- or PE-anti-β-globin (SantaCruz Biotech). The mAb and their corresponding isotypes were used at 1–2 mg ml^{-1}; optimal working dilutions were determined for individual antibodies. Frequencies of cells possessing the haemogenic and haematopoietic phenotypes were determined on live cells by 7AAD (Immunotech) exclusion, using FACSCalibur (Beckman Coulter), and analysis was performed using FlowJo 8.8.6 Software.

RT–PCR and qPCR. Total RNA was isolated using the Norgen RNA isolation kit. RNA was then subjected to cDNA synthesis using superscript III (Invitrogen). Quantitative PCR (qPCR) was performed using Platinum SYBR Green–UDP mix (Invitrogen). For the analysis of the sample, the threshold was set to the detection of Gus-B (β-glucuronidase)[46] and then normalized to internal control GAPDH. The base line for the experiment was set to the gene expression levels observed in fibroblasts. Given the expression of some of the genes within this starting population of fibroblasts, we felt that we should include the gene expression pattern for these cells. Hence, the data are represented as delta cycle threshold ($\Delta C(t)$) versus delta $\Delta C(t)$ ($\Delta\Delta C(t)$). (qPCR primer sequences are provided in Supplementary Table 3.)

Genomic DNA was isolated using the All In One isolation kit (Norgen). For integration studies, 150 ng genomic DNA was used per PCR reaction. PCR reactions were performed using 2X PCR Master Mix (Fermentas).

Affymetrix analysis. Total RNA was extracted from human dermal fibroblasts (2 replicates), puromycin selected day 4 OCT4 transduced fibroblasts (2 replicates) and sorted CD45$^+$ cells (2 replicates) using the Total RNA Purification Kit (Norgen). RNA integrity was assessed using Bioanalyser (Agilent Technologies). Sample labelling and hybridization to Human Gene 1.0 ST arrays (Affymetrix) were performed by the Ottawa Health Research Institute Microarray Core Facility (OHRI). Affymetrix data were extracted, normalized, and summarized with the robust multi-average (RMA) method implemented in the Affymetrix Expression Console. CEL files were imported into dChIP software[47] for data normalization, extraction of signal intensities and probe-level analysis. All the data from each individual sample are available on the NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE24621.

Chromatin immunoprecipitation. ChIP was performed as described previously[45]. Briefly, human pluripotent cells (H9 and iPSC1.2), human dermal fibroblast cells, puromycin selected day 4 OCT4 transduced cells, and sorted day 21 CD45$^+$ cells were cross-linked using 1% formaldehyde. Chromatin was digested in buffer containing 0.1% SDS to obtain fragments of approximately of 1,000 bp length. Sonicated DNA was subjected to immunoprecipitation using anti-OCT4 (ChIP quality antibody; Cell Signaling Technology) and anti rabbit IgG antibodies (Santacruz Biotechnology). Immunoprecipitated DNA was further reverse cross-linked, purified and subjected to qPCR analysis using UDG-Platinum SYBR Green mix (Invitrogen). The promoter specific ChIP primers are listed in Supplementary Table 4. To calculate the relative enrichment, signals observed in control antibody were subtracted from signals detected from the specific antibody; the resulting differences were divided by signals observed from one-fiftieth of the ChIP input material.

Megakaryocyte assay. To detect human megakaryocytes, the MegaCult-C Complete Kit with Cytokines (Stem Cell Technologies) was used. The derivation of megakaryocytes was done according to instructions included with the kit. The kit includes pre-screened components for optimal growth of megakaryocyte CFUs, such as thrombopoietin (TPO), Interleukin-3 (IL-3), IL-6, IL-11 and SCF, chamber slides for growth and antibodies for subsequent immunocytochemical staining. In short, 10,000 CD45$^+$ EPO treated cells were plated in the MegaCult medium containing the cocktail of growth factors mentioned above. The human CFU-Mks were detectable by day 10 to 15 and were subsequently fixed and stained according to protocol. Mk-specific antigen GPIIb/IIIa (CD41) linked to a secondary biotinylated antibody-alkaline phosphatase avidin conjugated detection system was used, where Mk-CFUs were red/pink in colour.

Cytospin. 1,000 CD45$^+$ OCT4 transduced cells were washed twice in cold 2% FBS in PBS and diluted in 500 µl of cold 1% FBS in PBS. The samples were loaded into the appropriate wells of the Cytospin centrifuge equipment. The samples were spun at 500 r.p.m. for 5 min to allow adherence to the slides. The slides were fixed with methanol for 1 min and allowed to dry for 30 min. Then slides were stained with Giemsa-Wright stain for 3 min, followed by 10 min in PBS and a quick wash in distilled water. The slides were allowed to dry overnight and mounted with mounting medium (Dako). Slides were viewed by an Olympus IX81 microscope. Blood cell typing/morphological criteria were confirmed by McMaster Pathology and Hematology.

Macrophage phagocytosis assay. Fluorescein (FITC) conjugated latex beads (Sigma) were used as particle tracers to analyse phagocytosis by monocytes derived from CD45$^+$Fibs^{OCT4} cells treated with IL-4 and M-CSF. To measure phagocytosis, 10 µl of packed beads suspended in 3% FBS in PBS was added to 10^6 cells in Teflon tubes. After incubation for 90 min at 37 °C, cells were washed three times with cold PBS containing 3% FBS and 0.1% EDTA to remove free beads. The cells were then labelled to detect expression of CD45 (APC-conjugated) CD45 mAb together with FITC-bead uptake, and analysed by flow cytometry using FACSCalibur (BD) or visualized on tissue culture quality slides (VWR) and viewed by an Olympus IX81 fluorescence microscope.

Xenotransplant assays. NOD/SCID IL2Rγc null adult mice (NSG) were sub-lethally irradiated with 325 rads 24 h before transplantation. 5.0×10^5 CD45$^+$ OCT-transduced (D37) or human dermal fibroblasts or human mobilized peripheral blood or human umbilical cord blood lineage depleted cells were transplanted by intrafemoral injection. After 10 weeks, animals were culled, and bone marrow (BM) from injected femur, contralateral bones and spleen was analysed for the presence of human cells by flow cytometry (FACSCalibur, Becton Dickinson), followed by data analysis using FlowJo 8.8.6 Software (Tree Star). Cells positive for HLA-A/B/C and CD45 were analysed for the expression of haematopoietic lineage specific markers, such as CD14. For secondary transplants, total engrafted bone marrow cells were transplanted intravenously (IV injection) in adult irradiated NSG mice, as described for primary transplants. Genomic DNA from engrafted cells was then analysed using conventional PCR by primers specific for the α-satellite of human chromosome 17: forward, 5′-GGGATAATTTCAGCTGACTAAACAG-3′; reverse, 5′-TTCCGTTTAGTTAGGTGCAGTTATC-3′.

Methylcellulose colony-forming assay. Cells were plated at 1,000 FACSAria II sorted (Becton Dickinson) CD45$^+$CD34$^+$ cells or 5,000 total cells for EPO treatments were seeded in 1 ml of Methocult GF H4434 (Stem Cell Technologies). Colonies were scored after 14 days of culture using standard morphological criteria and analysed using the FACSCalibur (Becton Dickinson) for haematopoietic surface markers. Collected events were analysed using FlowJo 8.8.6 Software (Tree

Star). For colony derivation from xenotransplant derived engrafted cells, the cells were first sorted based on HLA-A/B/C (BD Biosciences) followed by CD45 expression using a human specific anti-CD45 (BD Biosciences). The HLA-A/B/C and CD45 double positive cells were then plated at a density of 1,000 cell ml^{-1} in Methocult GF H4434. The colonies derived from engrafted cells were further analysed for haematopoietic surface markers using FACSAria II (Becton Dickinson). Collected events were analysed using FlowJo 8.8.6 Software (Tree Star).

Teratoma assay. The McMaster University Animal Care Council approved all procedures and protocols. Adult dermal fibroblasts, neonatal dermal (foreskin) fibroblasts, CD45$^+$ OCT4 transduced adult dermal fibroblasts, CD45$^+$ OCT4 transduced neonatal fibroblasts and iPSC 1.1 to 1.4 were treated with collagenase IV for 5–10 min, followed by collection and washing twice with saline and resuspended in saline. 500,000 cells per sample were injected intratesticularly into male NOD-SCID mice. Mice were killed 10–12 weeks after initial injection. Teratomas were extracted, embedded in paraffin and sectioned in 5-µm intervals followed by deparafinization in xylene and processing through a graded series of alcohol concentrations. Samples were stained with haematoxylin and eosin or OCT4 followed by dehydration and xylene treatment. Slides were mounted using Permount and imaged by scanning slides using Aperio Scan Scope; images were captured using Image Scope v9.0.19.1516 software. Tissue was also collected from a variety of organs including lung, spleen, liver, brain and kidney to investigate the presence of metastatic cells. Tissue typing was performed based on stringent histological and morphological criteria specific for each germ layer subtype. Mesoderm lineages, such as bone, were identified using presence of osteocytes and bone spicules (dark pink); cartilage was identified by the presence of chondrocytes and specific staining (light blue) of the extra cellular matrix. Endoderm lineages, such as intestinal lumens, were identified by the presence of goblet cells in the lumen epithelium. Ectoderm lineages, such as skin, were identified based on distinguishing cell layer morphologies (that is, stratified); brain or neural tube was identified based on specific histological criteria. The presence of the germ layers and tissue typing was confirmed by McMaster Pathology.

Statistical analysis. All tests were performed using InStat Version 3.0a statistical software (GraphPad Software). Descriptive statistics including mean and s.e.m. along with one-way ANOVAs, and two-tailed t-tests were used to determine significant differences. $P < 0.01$ was considered significant.

46. Oshima, A. *et al*. Cloning, sequencing, and expression of cDNA for human β-glucuronidase. *Proc. Natl Acad. Sci. USA* **84,** 685–689 (1987).
47. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98,** 31–36 (2001).